



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

객체 인식을 활용한
사진 관리 서비스의 사용자 경험 연구
- 구글 포토 내 검색 기능을 통한
되찾기(Retrieval)의 학습을 중심으로 -

2019 년 8 월

서울대학교 대학원

융합과학부 디지털정보융합전공

강 슬 기

객체 인식을 활용한
사진 관리 서비스의 사용자 경험 연구
- 구글 포토 내 검색 기능을 통한
되찾기(Retrieval)의 학습을 중심으로 -

지도 교수 이 중 식

이 논문을 공학석사 학위논문으로 제출함
2019 년 7 월

서울대학교 대학원
융합과학부 디지털정보융합전공
강 슬 기

강슬기의 공학석사 학위논문을 인준함
2019 년 7 월

위 원 장 _____ 이 재 현 (인)

부위원장 _____ 이 중 식 (인)

위 원 _____ 서 봉 원 (인)

초 록

지난 봄에 촬영한 벚꽃 사진이 찾고 싶을 때, 스마트폰의 사진첩에서 ‘벚꽃’을 검색해본다. 뒤이어 앨범 안에 있는 모든 벚꽃 사진이 나타난다. 이처럼, 스마트폰으로 촬영한 나의 사진도 이제 키워드를 통해 검색이 가능하다. 과거에는 앨범을 만들어서 관리하거나 수많은 사진의 홍수 속에서 스크롤 내려가며 브라우징했다면, 이제는 머리 속에 떠오르는 검색어를 활용해 원하는 사진을 찾을 수 있는 것이다. 이는 컴퓨터 비전 분야에서 객체 인식 기술이 비약적인 발전을 이루고, 비슷한 특징을 가지는 사진을 묶어낸 후 적합한 이름을 붙이는 데까지 가능해진 덕분이다. 그러나, 검색을 활용해보면 그 결과의 범위가 너무 넓어서 검색의 효과가 없거나, 사용자가 활용한 검색어에 부합하는 결과가 나오지 않는 경우도 있다. 거기에 정확지 않은 결과가 나타날 때도 있어 사진 검색을 활용하는 초기의 사용자에게 부정적인 경험을 남기게된다.

이에 본 연구는 객체 인식을 활용한 스마트폰 사진첩에서 나타나는 불편함의 지점을 밝혀내고, 사진 검색 경험을 개선하고자 하는 목적을 갖고 시작되었다. 객체 인식이 스마트폰 사진 검색과 같은 실생활에 적용되었을 때 나타나는 이러한 어색한 결과는 사진이라는 대상의 특성과 기기의 내부적 특성, 이로 인해 발생하는 기기와 사람 사이의 인식 차이로 인해 발생한다. 본문에서는 객체 인식의 내부적 특성을 문헌 연구와 사전 조사를 통해 수집한 후, 검색을 보다 효율적으로 할 수 있는 전략으로 일반화하였다. 이후 사용자에게 사진 검색의 경험 중 이러한 특성을 검색의 전략으로 활용할 수 있도록 팁(Tip)을 작성한 후 이를 실험의 참여자에게 제공하였다. 이후 16 명의 20-30 대를 대상으로 일주일동안 매일 여섯번의 사진 찾기 과업을 수행하는 실험을 통해 연구 자료를 수집하였다. 이를 통해 총 672 개의 검색 과업의 기록과 사용된 전략, 그리고 사용자가 형성한 전략을 수집했으며, 일주일간의 실험이 완료된 후에는 사후 설문 데이터를 얻을 수 있었다.

분석의 결과로, 먼저 전략을 통해 사용자가 사진 검색의 기능을 학습하는 모습을 확인할 수 있었다. 42 회의 검색 태스크가 누적됨에 따라 16 명의 평균 검색 소요 시간은 점차 감소하였으며, 이를 시간의 흐름에 따른 변화의 축에서 살펴보면 첫 번째 날에 비해 마지막 날의 평균 검색 소요 시간이 51 초에서 35 초로 31% 가량 감소했다. 검색의 평균 성공률 또한 일주일 동안 매일 진행된 42 회의 태스크에 걸쳐 약 11% 상승하는 모습을 보였다. 실험 참여자의 평균 검색 시도 횟수도 28% 감소하였다. 이를 통해 제공된 전략과 함께 사진 검색의 경험이 누적될수록 객체 인식 기술을 활용한 사진 검색에서 학습의 모습과 검색의 개선이 나타남을 확인할 수 있었다.

개별 사용자의 학습 형태에서는, 16 명 중 12 명의 참여자가 학습의 모습이 나타난 경우에 해당되었으며, 3 명은 학습이 이뤄지지 않은 모습을 보였다. 나머지 1 명은 학습에 영향을 받지 않는 모습을 나타내었다. 학습이 이뤄지지 않거나 영향을 받지 않는 경우, 초반의 검색에서 어떤 전략을 활용하고 검색 기능에 적응하는지에 따라 학습에서의 차이가 생기는 것으로 유추해볼 수 있었다.

마지막으로 제공된 전략 중 가장 많이 사용된 전략은 ‘상위어(상위 개념의 단어), 하위어(구체적인 단어)를 활용하여 정확한 명칭 사용하기’로 전체 검색의 44.35%이 해당 전략을 활용한 것으로 나타났다. 그 다음으로는 ‘,(콤마) 사용하기’, ‘색상과 같이 화면에서 드러나는 검색어 활용하기’, ‘인물(여자, 남자) 활용하기’가 뒤를 이었다. 사용자가 개별적으로 형성한 전략은 전체의 12.20%로 나타났으며, 시간이 흐름에 따라 그 사용의 차이는 나타나지 않았다. 사용자가 직접 제작한 전략으로는 ‘행정구역명의 활용’이 개별 전략의 39.47%를 차지했으며, 그 다음으로는 ‘자동 분류된 인물 활용’, 그리고 ‘건축물의 활용’이 주로 활용되었다. 형성된 전략 중 ‘구글 포토에 맞춘 단어 활용’은 과거의 경험을 통해 객체 인식의 특성을 인지하고, 이를 예측하여 검색어를 형성하는 것으로 나타났다. 즉, 객체

인식을 활용한 사진 검색의 경험을 누적할수록 그 특성에 대한 이해가 생기는 모습을 보여준다.

이상의 분석을 통해 객체 인식 기술이 스마트폰 사진첩 내에서 검색으로 활용될 때 사용자에게 어려움을 가져오는 지점에 대해 살펴본 후, 이를 보완할 수 있는 방안에 대한 간략한 제언을 덧붙였다.

본 연구는 객체 인식 기술이 스마트폰 사진첩에 적용되면서 나타난 사용의 어려움을 사용자의 관점에서 접근하였다. 더하여 HCI(Human Computer Interaction)의 측면에서, 기기의 관점을 통해 제작된 전략이 어떻게 사용자에게 수용되고 변형되는지 그 과정에 집중했다는 데 그 융합적 의의를 가진다. 더하여 인식의 정확도를 높이는데에만 집중되었던 객체 인식의 연구가 실제 사용자에게 제공되었을 때 발생하는 인터랙션을 실험을 통해 관찰을 시도했다는 데 의의를 갖는다. 마지막으로 사진이라는 매체의 활용성과 지속 가능한 활용을 위해 객체 인식을 활용할 수 있는 방식에 대한 논의를 펼쳤다는 데 의의가 있다.

주요어 : 머신러닝, 컴퓨터 비전, 객체 인식, 사진 검색, 사진 관리
학 번 : 2017-26435

목 차

제 1 장 서론.....	1
제 1 절 연구의 배경.....	1
제 2 절 연구의 목적.....	4
제 2 장 관련 연구.....	4
제 1 절 이미지 검색에서 객체 인식의 활용.....	8
제 2 절 기기와 인간의 관점 차이.....	11
제 3 절 사진의 특성과 스마트폰 사진 찾기.....	14
제 3 장 연구 문제.....	19
제 1 절 연구 문제의 설정.....	19
제 2 절 측정을 위한 개념 정의.....	21
제 4 장 연구 방법.....	24
제 1 절 연구 방법.....	24
제 2 절 실험 방법.....	25
제 5 장 연구 결과.....	41
제 1 절 객체 인식을 활용한 사진 찾기의 학습.....	41
제 2 절 사용자 별 학습의 형태와 그 특성.....	48
제 3 절 제공된 전략의 활용과 변형.....	52
제 6 장 연구 논의.....	63
제 1 절 연구에서 나타난 기기와 인간의 차이.....	63
제 7 장 결론 및 연구의 의의.....	67
제 1 절 연구의 요약.....	67
제 2 절 연구의 한계 및 제언.....	70
제 3 절 연구의 의의.....	72
참고문헌.....	74

Abstract.....	78
---------------	----

표 목차

[표 1] 최근까지 이미지 검색 연구의 변화 과정	10
[표 2] 제공된 사진 찾기 태스크의 예시	35
[표 3] 실험의 참여자 목록	36
[표 4] 태스크 수행 횟수에 따른 평균 소요 시간과 표준 편차	42
[표 5] 태스크 수행 일자의 누적에 따른 평균 소요 시간과 표준 편차	43
[표 6] 태스크 수행 일자의 누적에 따른 평균 성공 확률과 표준 편차	45
[표 7] 태스크 수행 일자의 누적에 따른 평균 검색 시도 횟수와 표준 편차	47
[표 8] 전체 태스크에서 사용된 전략과 그 비율	53
[표 9] 사용된 전략 중 검색의 실패 비율	54
[표 10] 시도 횟수의 누적에 따른 ‘나만의 전략’ 활용	56
[표 11] 참여자가 형성한 전략과 그 비율	57

그림 목차

[그림 1] 구글 포토에서 객체 인식 기술을 통해 분류한 카테고리의 예시	2
[그림 2] 선행 연구 탐색의 범주	8
[그림 3] 가이드 문서의 예시	26
[그림 4] 제공된 팁(Tip)의 예시	29
[그림 5] 챗봇의 메뉴에서 확인할 수 있는 팁의 예시	30
[그림 6] 실험에 활용된 챗봇의 예시	31
[그림 7] 진행 과정에 따른 실험 진행자와 참여자의 역할	37
[그림 8] 수집된 데이터의 예시	40
[그림 9] 태스크 누적 횟수에 따른 소요 시간의 변화	42
[그림 10] 수행 일자의 누적에 따른 소요 시간의 변화	44
[그림 11] 태스크 누적 횟수에 따른 평균 성공 확률의 변화	45
[그림 12] 수행 일자의 누적에 따른 평균 성공 확률의 변화	46
[그림 13] 태스크의 누적 횟수에 따른 평균 검색 시도 횟수의 변화	47
[그림 14] 수행 일자의 누적에 따른 평균 검색 시도 횟수의 변화	48
[그림 15] 검색 소요 시간이 감소하는 경우의 예	49
[그림 16] 검색 소요 시간이 증가하는 경우의 예	50
[그림 17] 학습 추이를 확인하기 어려운 경우의 예	52
[그림 18] 수행 일자에 따른 전략 사용의 변화	56

제 1 장 서 론

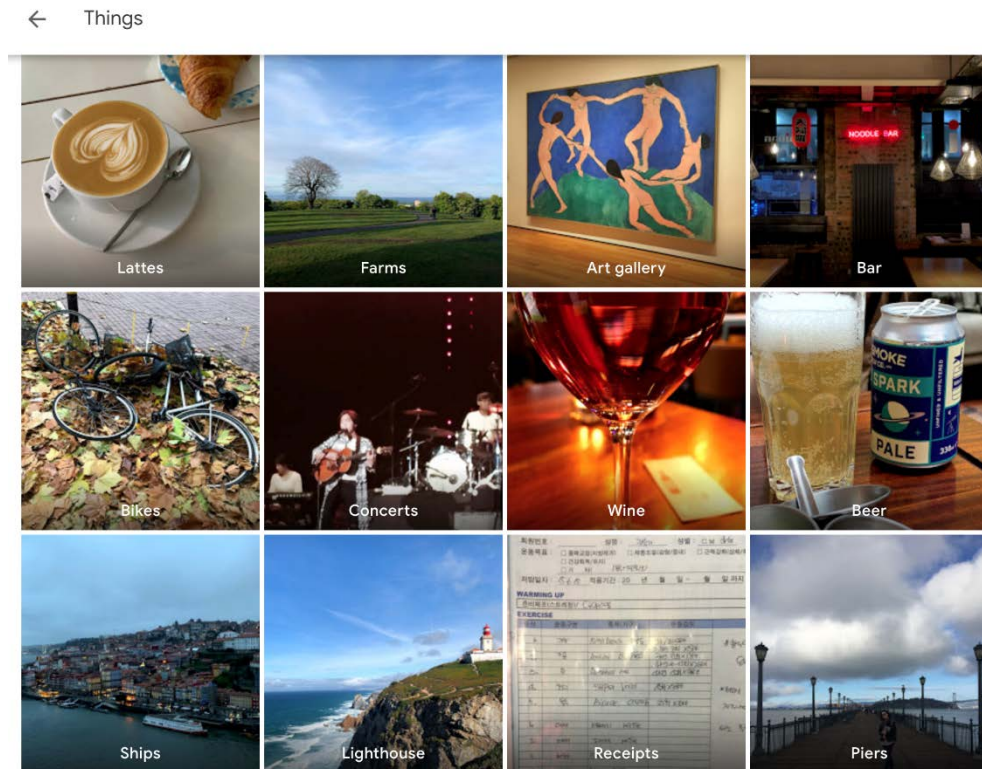
제 1 절 연구의 배경

스마트폰 사진첩은 점차 복잡해지고 있다. 사람들은 일상 속에서 인상적인 순간에 손쉽게 사진을 촬영하고, 의미 있는 한 장면을 남기기 위해 연속으로 수많은 사진을 촬영한다. 과거의 사진은 가족이나 친구, 혹은 생일과 같은 이벤트를 중심으로 촬영되었다. 반면 최근의 스마트폰은 휴대하기도 쉽고 촬영한 사진을 다른 사람들과 쉽게 공유할 수 있다. 이러한 스마트폰의 특징에 따라 다양한 유형의 사진이 스마트폰 사진첩에 촬영되고 저장되는 모습을 확인할 수 있다. 예를 들어 사진첩 내에서 인터넷 이미지, 스크린 샷, 영수증, 여권과 같은 다양한 사진의 형태가 등장하는 것이다 (Broekhuijsen et al., 2017).

이처럼 폭발적인 사진의 양의 증가와 다양해진 사진의 형태는 사진 관리에 어려움을 더하였다 (Sarvas et al., 2011). 스마트폰 사진첩 내부의 사진은 점차 쌓여나갔고, 관리는 뒷전으로 물러났다. 사람들은 사진 관리에 대한 동기와 그 관리 방식을 잊어버리기 시작했다. 넘쳐나는 스마트폰 내 사진을 관리하려면 더 많은 노력이 필요해졌지만, 사람들은 이를 어렵고 불편한 작업으로 생각하여 시도하지 않게 된 것이다 (Van House, 2011). 따라서 정작 사진을 둘러보거나, 특정 사진을 찾으려 스마트폰 사진첩을 들어가면 사막에서 바늘을 찾는 것과 같이 수많은 사진의 나열 속에서 해매는 상황이 발생해버리고 만다.

사진을 효율적으로 관리하기 위한 시도가 이어졌지만, 예부터 사진 관리는 처치 곤란의 대상이었다. 늘어나는 사진의 양에 반해 더 이상 관리에 소홀해지는 사람들로 인해 사진은 점차 되찾기 어려운 대상이 그러나 최근 컴퓨터 비전 영역에서 기계 학습(Machine Learning)기술이 발전한 후, 객체 인식(Object Recognition) 기술이 사진첩에서 ‘공동

관리자”의 역할을 도맡아 사진을 관리하기 시작하였다. 컴퓨터 비전의 패턴 인식이 고도화됨에 따라 기존에 학습한 객체를 대상으로 사진첩의 사진의 화면에 나타나는 객체나 장면을 읽어내고, 비슷한 사진을 찾아 자신의 방식대로 분류한 후 이를 대표할만한 라벨을 부여하기 시작한 것이다. 이에 따라 사용자의 사진첩 내에서 비슷한 사진들이 한데 묶여 사람들에게 제공되었다 [그림 1]. 나아가 사진에서 등장하는 사물이나 ‘미팅(meeting)’과 같이 화면에서 드러나는 이벤트를 검색어로 입력하면, 해당 단어로 분류된 사진들이 결과로 제공된다. 즉, 객체 인식 기술이 혼잡한 사진첩을 관리해주는데 까지 그 활용 영역을 넓힌 것이다.



[그림 1] 구글 포토에서 객체 인식 기술을 통해 분류한 카테고리의 예시

기존에는 사진 촬영 시 자동으로 기록되는 GPS(Global Positioning System) 정보, 날짜와 같은 메타 데이터가 단순히 사진을 분류하는데 활용되었다면, 이제는 장소와 날짜, 계절까지 검색어로 활용할 수 있다. 나아가 사진의

화면에서 보여지는 사물과 활동을 검색어로 사용하여 원하는 사진에 효율적으로 접근할 수 있는 가능성이 열렸다. 즉, 기존의 브라우징을 통한 사진 찾기 방법에 비교했을 때, 사진을 ‘검색’을 활용하여 찾을 수 있다면 훨씬 효율적인 되찾기가 가능해지는 것이다. 이는 사진을 되찾는 방법에 효과적으로 활용될 것으로 기대해볼 수 있다.

그러나 현재 구글 포토에서 검색어를 활용하여 검색을 하면, 많은 경우 결과가 제대로 나오지 않는다. 혹은 여러 번의 시도 후에 어떤 검색어를 활용했을 때 구글 포토가 인식할 수 있는지를 파악할 수 있다. 반복적인 시도를 통해 사진 검색 기능을 학습한 후에야 겨우 원하는 사진에 더 빠르게 접근할 수 있게 되는 것이다. 혹은 검색어를 활용하여 결과는 얻었지만, 제대로 된 분류가 이뤄지지 않아 원하는 사진은 찾지 못하는 경우도 더러 있다. 달리 말하면 기술의 발전으로 사진을 질의어를 통해 되찾을 수 있는 검색 기능이 가능해졌지만, 그에 반해 사용성과 검색의 효율은 매우 떨어지는 상황인 것이다. 이에 따라 사진첩에서의 검색 기능이나 사물의 분류는 한번씩 알 수 없는 결과를 내놓아 재미를 가져다 주는 기능이 되어버렸다. 현실적인 활용 빈도 또한 매우 낮은 실정이다. 결국 머신 러닝과 객체 인식의 기술을 사진 되찾기(Photo Retrieval)의 분야에 접목되었을 때, 실생활에서 이를 활용하는 초기의 사용자들은 기대치에 미치지 못하는 결과에 결국 부정적인 인식을 가지게 되고, 지속적인 사용으로 이어지지 않는 것이다. 기술은 적용되었지만, 아직 실생활에 적용되기에는 기술이 성숙되지 않았거나, 기술의 결여를 초기 사용자 경험이 보조하지 못하는 경우인 것이다.

그러나 검색의 대상이 되는 스마트폰의 사진은 일반적인 이미지와는 달리 개인의 기억을 비추는 거울이다. 나아가 과거의 사건과 사물, 어떤 현상이 존재했음을 증명해주는 매체이기도 하다. 특히 스마트폰을 통해 촬영된 사진은 일상의 누적이자 개인의 생활과 경험을 비춰주는 창구로 해석될 수 있다. 사진을 검색하는 행위는 이렇듯 스마트폰을 통해 촬영된 사진의 의미를 극대화하기 위한 시도 중 하나로 제안된 것이다. 그러나

실제 사용자가 머리속의 사진을 검색어를 통해 찾는 일은 아직까지 어려운 일이다. 즉, 아직 사용자가 기억 속에 떠올리는 사진에 빠르고 쉽게 접근할 수 없으며, 객체 인식 기술을 활용하더라도 모든 종류의 사진을 효율적으로 되찾는 일은 해결될 수 없는 난제처럼 남게 된다. 이에 본 연구는 스마트폰 내부에 깊이 잠들어있는 개인의 사진을 쉽고 빠르게 되찾을 수 있는 방법에 대한 궁금증으로부터 해결 방안을 모색하기 위해 끼우는 첫 단추로서 시작되었다.

객체 인식이 활용된 사진 검색 기능의 활용을 증진시키려면, 현재의 사용에서 나타나는 어색함과 부정확함이 개선되어야 한다. 그러나, 객체 인식의 정확도를 높이는 일은 현재 컴퓨터 비전 분야에서 활발히 진행중에 있다. 본 연구는 객체 인식 기술이 가지는 근본적인 문제를 먼저 문헌을 통해 살펴봄으로써, 사용자가 기기 내부의 특성을 이해하거나 우회하여 활용할 수 있는 방안에 대해 모색하고자 했다. 이는 사용자에게 사진을 잘 찾기 위한 전략으로 제작되어 그 효과를 파악할 예정으로 만들어졌다. 나아가 현재 사진 검색의 기능이 적용되어 사용되는 과정 속에서, 머신러닝의 특성이 드러남에도 불구하고 사용자가 객체 인식을 활용한 사진의 검색에 적응하는지 그 여부를 일정 기간 동안의 학습 과정을 통해 살펴본다. 결과적으로는 객체 인식이 적용된 사진 검색에서 사용자가 갖게 되는 문제 지점을 파악하고, 이를 극복하기 위해서는 사용자의 관점에서 어떤 지점을 해결해야하는지 살펴보고자 한다.

제 2 절 연구의 목적

본 연구는 최근 객체 인식을 활용한 스마트폰 사진의 검색이 가지는 부정확하고 비효율적인 사용을 극복하고자 하는 필요에서부터 시작되었다. 이를 해결하기 위해 사진 검색을 활용하고 이에 적응해나가는 모습을 사용자의 관점에서 관찰하고, 사용자에게 기기의 관점을 이해할 수 있도록 제공하는 전략의 사용을 확인해본다. 여기서 사용자의 관점은 사진 검색을

사용하는 행태를 뜻하며, 구체적으로는 객체 인식을 활용한 사진 검색 기능을 처음 사용하는 사람을 대상으로 그 학습 과정을 의미한다. 기기의 관점은 시스템 내부에서 작동하는 객체 인식과 머신 러닝의 특성을 담은 전략으로 학습의 과정 중에 이를 제공함으로써 사용자가 이를 활용 및 변형하는 행태를 관찰한다.

특히, 객체 인식을 활용한 사진 검색을 처음 사용하는 사용자를 관찰의 대상으로 하여 해당 사용자가 검색 기능에 적응하는 과정을 살펴볼 필요가 있다. 통상적으로는 검색의 전략을 수집하고 이를 분석하기 위해 능숙한 사용자를 대상으로 하여 효율적인 검색의 행태를 관찰하였다. 그러나, 객체 인식 기술이 개인 사진을 검색하는데 적용되어 사용자에게 서비스로 제공되는 경우는 비교적 최근에 등장한 행태이다. 이에 따라 숙련된 사용자를 관찰하기보다는 구글 포토에서 검색 기능을 처음 활용해보는 사용자로 한정지어 검색 기능을 학습해 나가는 과정을 살펴보고, 현재 시스템이 가지는 어려움을 극복하기 위해 자신만의 전략을 만들어나가는 과정을 살펴봄으로써 추후 개인 사진 검색의 개선 가능성을 모색하고자 하였다.

객체 인식을 활용한 사진 검색의 내부적인 특성을 사용자에게 전략의 형태로 제공한 후 사용자가 학습해 나가는 과정을 살펴본다. 이를 통해 현재의 사진 검색이 가지는 어려움은 어떤 이유에서 생겨나는지, 극복을 해결할 수 있는 부분이 어디에 위치해 있는지 파악한다. 이를 통해 사진 검색 경험을 증진시킬 수 있는 첫 출발이 되는데 기여하고자 한다. 따라서 사진 검색 기능을 사용하고 제공되는 전략을 통해 검색 방식을 학습하는 사용자의 관점과, 객체 인식과 인간이 가지는 인식과 분류 체계의 차이, 그리고 개인의 사진이라는 독특한 특성을 바탕으로 다음과 같은 연구 목적을 가진다.

1. 사용자가 객체 인식을 활용한 사진 검색을 학습할 수 있는지 확인한다.

2. 사용자가 사진 검색을 위해 제공되는 전략을 얼마나 활용하는지, 검색 경험에서 어려움을 극복해가는 개인의 전략은 어떻게 나타나는지 확인한다.
3. 객체 인식을 활용한 사진 검색의 어려움이 발생하는 지점과 그 원인을 파악한다.

제 2 장 관련 연구

본 연구에서 관찰하고자 하는 사진 검색 (Photo Retrieval) 행동은 최근 객체 인식 기술을 활용하여 스마트폰 사진첩의 사진을 읽어내고, 감지한 부분에 적합한 라벨을 부여한 후 같은 라벨이 붙은 사진들을 한데 묶어 사진 관리(Photo Management)와 찾기(Retrieval)를 편리하게 만들려는 시도의 결과이다. 먼저 사용자의 사진첩에 누적되고 쌓여가는 사진들을 자동적으로 분류하고 라벨을 붙인다. 그 후 분류의 결과를 사용자가 검색 기능을 통해 활용하도록 만드는 시도가 곧 스마트폰에서 일어나는 객체 인식을 활용한 사진 검색인 것이다. 여기에는 사진의 검색이 가능하도록 기여한 기술적 배경과 일반적인 이미지가 아닌 사진이기 때문에 나타나는 특성 등 연관된 개념들이 복잡하게 얹혀있다. 따라서 해당 개념들을 하나씩 짚어보고 이에 대한 선행 연구를 먼저 탐색할 필요가 있다.

관찰의 초점은 객체 인식 기술이 접목된 검색 기능에 대한 학습과 검색의 어려움을 극복해나가는 전략의 사용이지만, 그 전에 왜 이러한 어려움이 발생하는지에 대한 이유를 살펴보고, 이후 데이터의 해석에 활용하기 위해 객체 인식과 인간의 인식 사이에 나타나는 괴리를 먼저 살펴볼 필요가 있다. 일례로 화면에서 드러나는 피쳐(feature)를 기반으로 하여 사진을 읽어내고, 분류하는 객체 인식의 기본적인 인식 과정과 기존 지식을 바탕으로 사물을 읽고, 알고 있는 개념 내에서 이름을 붙이는 인간 사이에는 필연적인 차이가 발생할 수밖에 없는 것이다. 더불어 단순히 일러스트레이션 혹은 발표 자료를 만들기 위한 이미지를 검색하는 행동은 일반적인 정보 검색 모델로도 설명할 수 있다. 그러나, 본 연구에서는 스마트폰에서 촬영된 사진이라는 특수한 맥락을 가진 대상을 다룬다는 점에서 일반적인 이미지와는 차이를 가지며, 머신 러닝을 통한 자동화가 관리의 과정 중에 사용된다는 점에서 그 특징을 가진다.

객체 인식, 그리고 인식의 대상이 되는 스마트폰의 사진, 그리고 사람과 기기의 차이점에 대한 이해를 위해 선행 연구는 [그림 2]와 같이 세 가지 범주에서 이루어졌다. 첫 번째로는 본 연구가 등장하게 된 배경이 되는 컴퓨터 비전의 발전과 이를 활용하여 이미지 검색에 활용하는데 이어지는 흐름을 살펴보았다. 두 번째로는 객체를 인식하는 사람과 기기 간의 차이에 대해 이해하였다. 세 번째로는 객체 인식의 대상이 되는 사진에 대한 다양한 관점을 살펴보고, 스마트폰 사진의 특성과 그 관리 및 되찾기에 대한 현 상태를 파악하고자 하였다.

[그림 2] 선행 연구 탐색의 범주

1. 컴퓨터 비전의 발전과 객체 인식

이미지를 인식하고 분류하는 일이 가능해졌다. 구글에서는 이미지의 개별 요소를 인식할 수 있도록 심층신경망을 구축하였는데, 이 심층신경망은 스스로 특정한 개념을 만들어낸 후 이미지를 분류하는데 성공하였다 (Le, 2011). 예를 들어, 이미지 내에서 고양이의 얼굴을 바운딩 박스로 지정하여 해당 부분이 ‘고양이’임을 학습시키는 것에서 나아가, 심층신경망을 통해 동그라미 위에 세모가 두개 있고, 동그라미의 안에는 작은 원이 두개 있다는 패턴을 스스로 파악해내는 것이다. 여기에 학습한 이미지 형태의 패턴이 어떻게 불러야 할지는 모르지만, ‘고양이의 얼굴’이라는 공통된 개념을 가진다는 사실을 만들어낸다.

이전까지는 알고리즘을 통해 형성해낸 분류에 대해 ‘고양이’라는 인간이 인지하는 개념의 이름을 달아준 후 학습을 할 수 있도록 만드는 일은 사람의 것이었다. 하지만 구글의 연구가 기존 연구에서 나아간 점은, 인식된 분류 패턴인 ‘고양이의 얼굴’이라는 개념에 사람이 사용하는 표현을 학습하여 ‘고양이 얼굴’이라는 텍스트 태그를 연결 지어준다는 것이다 (오세욱, 2017). 인식한 개념에 대해 어떻게 명칭을 붙여야 할지 스스로 판단이 가능해진 점에서 비약적인 발전을 보이고 있다. 컴퓨터 비전 기술은 자율 주행 시 객체를 인식하는 데에서부터 사람의 얼굴에 나타나는 감정 인식 등 다양한 분야에서 활용될 수 있다. 이미지 검색(Image Retrieval)의 분야에서도 객체 인식 기술은 기여하는 바가 크다.

2. 객체 인식과 이미지 검색

이미지를 찾기 위한 방법에 대한 연구는 오래전부터 진행되어 왔으며, 대표적으로 텍스트 기반의 이미지 검색(Text-Based Image Retrieval, 이하 TBIR) 방식과 내용 기반의 이미지 검색(Content-Based Image Retrieval, 이하 CBIR) 방식이 있다. TBIR의 경우, 이미지의 내용과 연관된 태그나 메타데이터를 통해 연관 이미지를 불러오게 된다. 이 과정에서 텍스트는 이미지를 검색하기 위한 검색어로 사용된다. 다만, 과거에는 이미지에 맞는 주석을 사람이 직접 손으로 작성하고, 맵핑하는 형태였다. 따라서 시간과

비용이 많이 든다는 치명적인 단점을 가졌다. 이와는 달리 이미지 내에서 시각적 특징을 도출하여 유사한 이미지를 찾는 CBIR 의 경우 비교적 효율적인 방식으로 대두되었다. 그러나 이미지에서 드러나는 단순히 색상, 질감과 같은 낮은 수준의 키워드와 그 안에 담긴 의미와 같은 높은 수준의 키워드 사이의 괴리(Semantic Gap)은 현재까지 해결되지 못한 채 남겨진 상황이다 (Yang, 2012). 특히 대부분의 CBIR 은 낮은 수준의 키워드나 얼굴, 차, 실내, 실외와 같은 특정 개념을 인식하는데 초점이 맞춰져 있고, 시스템의 최종 사용자나 이미지를 인출하는데 연관된 사람의 특성을 무시한다는 비판을 받고 있다 (Jaimes, 2006).

그러나 2010 년 이후, 딥러닝 기술은 이미지에 의미적으로 연관된 주석(Annotation)을 달거나, 나아가 검색 과정에도 적극적으로 사용되기 시작했다 (Bhagat & Choudhary, 2018). 이는 CBIR 의 한계를 대체할 수 있는 방법으로 재조명 되고있으나, 지도 학습(Supervised Learning)의 경우 라벨이 되어있는 수많은 학습 데이터가 필요하다는 한계에 이내 부딪히게 된다. 이에 따라 비지도학습(Unsupervised Learning)을 기반으로 하는 반-지도 학습(Semi-supervised Learning) 연구가 진행되기 시작하였지만, 이 또한 빠르게 증가하는 노이즈 이미지들과 그에 따라 정돈되지 않은 텍스트 데이터 세트를 통해 어떻게 학습하는지에 대한 어려움에 직면하고 있다 [표 1].

[표 1] 최근까지 이미지 검색 연구의 변화 과정

	초기 단계	중기 단계	최근
특징 추출 방식	손으로 선정	시각적 요소와 텍스트의 관계	시각적 요소와 텍스트의 관계, 딥러닝 피쳐(Features)
학습 방법	지도학습	지도학습	반-지도학습 (비지도학습)
주요 검색 방식	CBIR	CBIR	TBIR

문제점	의미적 괴리	이미지 기반 검색어 생성	정돈되지 않은 주석 텍스트를 통한 학습
-----	--------	------------------	--------------------------

3. 이미지 검색과 디지털 에크프라시스

에크프라시스(Ekphrasis)는 고대 그리스 시대 수사학의 한 과목으로, ‘시각적 표상이나 장면에 대한 언어적 표상’을 뜻한다. 이재현(이재현, 2009)은 해당 개념을 기술 발전에 따른 멀티미디어와 웹 검색 등에 접목하여 ‘디지털 에크프라시스’를 정의하였다. 그는 과거에는 그림이 텍스트에 비해 우위를 차지한 반면, 멀티미디어의 시대에서는 텍스트가 ‘호명 장치’로서 그림에 비해 우위를 차지하고 있다고 역설하였다. 검색 엔진에서 메타데이터나 태그는 콘텐츠를 불러내는 역할을 하게 되고, 이로써 언어가 멀티미디어의 콘텐츠에 앞서는 ‘에크프라시스의 희망’, 즉 TBIR 을 강조하게 되는 것이다. 더하여 그는 ‘에크프라시스의 불가피성’이라는 표현과 함께 TBIR 의 사용을 강조하고 있는데, 이는 앞서 살펴본 이미지 검색 기술의 발전 과정에서 보았듯 CBIR 에서 TBIR 의 활용으로의 전환을 통해 확인할 수 있다.

본 연구에서는 사진이라는 멀티미디어를 질의어를 통해 검색한다는 점에서 사진을 ‘호명’하여 그 존재를 확인하는데 언어적 표상의 양식을 띤다고 할 수 있으며, 수많은 사진 속에서 보다 ‘효율적’이고 ‘정확하게’ 찾고자 하는 에크프라시스의 욕망에서 그 의미를 찾을 수 있다.

제 2 절 기기와 인간의 관점 차이

객체 인식 연구의 분야에서 대량의 데이터를 바탕으로 이미지와 그 개념을 학습시키는 시도를 한 ImageNet 프로젝트는 실제로 분류의 정확도가 인간보다 높아지는데 큰 기여를 했다. 실제로 2010년 이후 이미지 레이블링에 대한 오류율이 28.5%에서 2.5% 미만으로 감소하였으며, 2015년

이후에는 인공지능을 활용한 객체 인식의 정확도가 인간의 수행 능력을 추월하는 데까지 성공했다^①.

실제로 ImageNet 연구팀에서 진행한 연구에 따르면, 컴퓨터가 사람에 비해 더 많은 오류를 내는 경우는 먼저 배경에서 선글라스를 쓰고 있는 사람과 같이 작거나 얇은 물체가 등장하는 경우, 과도한 필터의 적용으로 인해 화면 색상의 왜곡이 발생하여 인식이 어려운 경우, 3D 이미지 혹은 동상과 같이 추상적인 형태를 가지는 경우가 있다 (Russakovsky et al., 2015). 반면 사람이 컴퓨터에 비해 인식과 분류의 정확도가 떨어지는 경우는 포메라니안 혹은 그와 동급의 개 품종을 인지하고, 적합한 이름을 붙이는 상황에 주로 발생한다. 이는 사람이 수백 가지의 개 품종을 모두 알고 있지 못하기 때문에 생기는 경우이며, 나아가 사람은 기존에 알지 못하는 대상에 대해서는 이를 호명할 수도, 화면에서 인식하기도 어려운 상황이 발생한다. 이처럼 현재 객체 인식에서 발생하는 간극은 기술이 실제 서비스로 적용될 경우에도 사용자에게 괴리감을 제공할 수 있다.

1. 사람이 객체를 인식하는 방식

인지심리학에서 인간이 대상을 인식하고 이를 뇌 속에서 처리하는 과정을 이론화한 정보 처리 이론(Information Processing Theory)에 의하면, 사람이 어떤 대상을 인식하고, 이를 기억할 수 있는 저장소까지 다르는데까지는 크게 세가지 단계가 존재한다. 먼저 대상이 눈을 통해 상으로 맺히게 되면, 사람은 선택적으로 대상에 주의를 가지게 된다. 이후 선택된 대상에 대한 해석과 범주화 과정을 거치게 된다. 이 과정을 통해 대상을 기존의 기억에 있는 유사한 대상과 연결하거나 결합하기 위해, 혹은 새로운 정보로 저장하기 위해 인식된 정보를 단순화하여 뇌 속의 장기 기억 저장소에 보관하는 과정을 거치게 된다 (조수진, 2012).

^① AI Index 2018 연례보고서 <http://cdn.aiindex.org/2018/AI+Index+2018+KOREAN.pdf>

사람이 그림을 눈으로 읽어낼 때에는 화면에서 기억되는 정보 요소는 크게 3 가지로 나타난다 (Mandler & Parker, 1976). 가장 먼저 화면에 드러난 대상과 대상의 물리적 특성에 대한 묘사나 세부적인 정보, 그리고 대상 사이의 공간적 거리를 파악한 후, 그 외에도 장면의 공간적 구성, 질감과 같은 낮은 단계의 특성 등을 기억하게 된다. 대상을 인식한 후, 사람은 머리 속에서 인식한 대상에 대해 적합한 이름을 부여하게 된다. 이름을 부를 때에도 사람은 대상을 지칭하는 개념 중 자연스럽게 특정 위계의 단어를 선택하게 된다. 대상을 부르는 개념 사이의 관계를 밝힌 로슈(Rosch, 1978)의 연구에 따르면, 대상을 분류할 때에도 크게 세가지의 분류가 가능한데, 기본 단계(Basic-level), 상위 단계 (Superordinate-level), 그리고 종속적 관계(Subordinate-level)가 바로 그것이다. 나아가 이후의 연구에서는 사람이 인지하고 사용하는 분류의 방식을 진입 단계(Entry-level)로 이름 지어 사용하고 있는 모습을 확인할 수 있다 (Jolicoeur et al., 1984). 최근 객체 인식 연구 분야에서도 인식의 결과에 어떤 라벨을 붙여야 하는지에 대한 고민이 이어지고 있는데, 그 결과에 따르면 사람들이 가장 자연스럽게 활용하는 단계는 진입 단계(Entry-level)인 것으로 나타났다 (Ordonez et al., 2016). 그러나, 진입 단계의 단어는 그 위계가 객체에 따라 달라질 수 있다. 일례로 미국 지빠귀(American Robin)과 펭귄(Penguin)의 기본 단계의 단어는 모두 새(bird)이나, 사람들로 하여금 이름을 부르도록 요청하는 경우 전자의 진입 단계 단어는 새(bird)로, 후자는 펭귄(penguin)으로 사용된다. 이는 그 객체가 가지는 특성에 따라 달라지게 된다.

2. 기기가 객체를 인식하는 방식

ImageNet 연구팀의 목표는 컴퓨터가 사람처럼 화면을 읽고, 그 의미를 알 수 있도록 만드는 것이었다. 연구진이 선택한 방식은 10 만장 이상의 사진에 대해 사람들로 하여금 개별 사진이 어떤 사물을 담고 있는지, 어떤 사진인지에 대한 레이블을 지정한 후, 이 이미지들을 머신 러닝의 인공 신경망을 통해 학습시키는 것이었다 (Russakovsky et al., 2015). 이를 통해 새로운 이미지가 읽혔을 때, 과거의 학습을 바탕으로 가장 비슷한 특징을

가진 객체로 분류하여 같은 라벨을 부여하는 것이다. 그 차이를 읽어내는 것이 고양이와 개 사이의 차이점을 파악하는 데에서 나아가 같은 개라도 다른 품종을 구분해내는데 까지 연구가 진행중에 있다.

객체 인식이 진행되기 위해 가장 중요한 과정은 화면의 피쳐(feature)를 추출해내는 과정인데, 이미지 내의 특성인 피쳐를 선정하는 방식에는 크게 두 가지가 있다 (Zhou et al., 2017). 먼저 색상이나 모양, 질감이나 구조를 전반적으로 하나의 파악해내는 방식(Global features)이다. 다음으로는 일부 영역을 바운딩 박스로 지정하여 국소적으로 특성을 읽어내는 방식(Local features)이 있다. 전자는 화면 내에서 혼잡한 배경과 인식해야하는 대상이 같이 있을 경우에는 감지가 어렵다는 단점을 가지고 있다. 후자의 방식은 두 단계를 거쳐 피쳐의 추출이 이뤄지는데, 먼저 특정 포인트가 되는 지점을 선정한 후 해당 부분에 대한 묘사를 진행하여 정보를 얻어내는 방식이다. 보통 하나의 이미지에서 적개는 몇 백개 혹은 몇 천개의 로컬 피쳐가 추출될 수 있으나, 전반적으로 피쳐를 추출하는 글로벌 방식의 경우에는 이미지의 부분적 정보를 가질 수 없다. 이처럼 컴퓨터가 이미지를 읽어내는 방식은 화면 내에서 드러나는 특징을 기반으로 해당 특성을 가지면 그와 가장 유사한 피쳐를 가진 객체의 이름을 가지는 단계를 걸쳐 진행된다. 최근에 들어서 주목 기제(Attention Mechanism)을 활용하여 화면의 전체를 읽는 것이 아닌, 일부분에 초점을 맞추고 필요한 정보를 얻을 수 있는 시도들이 등장하고 있는데, 현재까지의 연구에 따르면 사람이 사진을 읽을 때 초점을 맞추는 영역과 기기가 초점을 맞추는 부분에는 차이가 존재한다 (Das et al., 2017).

제 3 절 사진의 특성과 스마트폰 사진 찾기

사진은 예부터 다각적인 관점에서 해석이 가능한 대상으로, 여러 분야의 지식인들에게 논의 대상이 되어왔다. 본문에서 다루게 될 스마트폰 사진에 대해 논하기 전에 고전적으로 논의되어왔던 사진에 대한 다양한

해석을 살펴본 후, 일반적인 이미지와 달리 촬영된 사진이 객체 인식에 적용되었을 때 차이를 가져오는 지점에 대해 알아본다. 본 절의 목적은 스마트폰이 가져온 변화와 맞물려 사진이 가지는 고유한 특성을 파악하여 개인의 사진을 객체 인식하는데 있어 생기는 차이점과 그 의미를 읽어내는 시각을 가질 수 있는 출발점을 다지는 데 있다.

1. 사진에 대한 해석

수전 손택은 〈사진에 관하여(1973)〉에서 사진을 세계를 전유할 수 있게 해주는 기법이자, 하나뿐인 자아의 표현으로 기술한다. 사진을 촬영하는 촬영자는 자신만의 방식으로 세상을 바라보고, 그 기준에 맞춰 현실을 해석하는 것이다. 사진을 촬영하는 행위에 대해 ‘이것(대상)을 보는 행위가 기록으로 남길 만한 가치가 있다고 결정했다’고 표현한 존 버거의 〈사진의 이해(1980)〉에서도 볼 수 있듯, 사진은 촬영자의 시각과 관점이 매우 큰 영향을 미친다. 즉, 사진은 보는 사람에 따라 그 관점이 다르기에 같은 대상이라도 다양한 해석을 가질 수 있게 된다.

사진은 사건 혹은 사물이 ‘거기에 있었음’을 증명하는 매체로 해석되기도 한다. 롤랑 바르트는 사진에 대해 촬영자가 현재 보고 있는 것이 실제로 ‘존재 했음’을 밝히는데 그 의미가 있음을 기술하였다. 나아가 사진사는 순간을 담아냄으로써 다른 사람에게 정보를 제공하고, 자신의 사진을 통해 다른 사람의 기억 속에서 잊혀지지 않도록 하는데 그 의도를 갖고 사진을 촬영한다 (Flusser, 1983). 그러나, 시스템 혹은 머신 러닝(Machine Learning)이 사진 촬영자의 의도를 읽어내는 것은 현재로서는 불가능하다. 본 연구의 관점에서 바라볼 때, 결국 머신러닝의 시야에 들어오지 못한 사진들은 잊혀지게 되고, ‘호명되지 못한, 읽히지 못한 사진들’은 점차 혼잡한 사진첩 내부에서 헤어나오지 못한 채 결국 사람들의 기억 속에서도 사라지는 결과를 초래하게 된다.

바르트는 그의 저서 <밝은 방(Camera Lucida)>를 통해 사진을 구성하는 두가지 차원을 스튜디오(studium)와 펀크툼(punctum)으로 대비하여 설명하였다 (Barthes, 1981). 스튜디오는 사진의 프레임 안에서 사람들이 익숙하고 친근하게 지각하는 영역을 뜻하는데, 이는 개인의 지식과 문화에 의해 영향을 받는다. 관람객은 사진에서 읽어낼 수 있는 코드를 학습하게 되고, 이를 통해 사진은 하나의 정보로써 기능하게 되는 것이다. 이와 반대로 펀크툼은 사진을 관람하는 대상에게 “찌르는 듯 한” 느낌을 제공하는, 규정되기 어려운 요소이다. 이러한 점에서 볼 때 객체 인식은 단순히 화면에서 보이는 대상을 보편적인 언어로 해석해낸다는 점에서 스튜디오의 특성을 지닌 것으로 해석할 수 있다. 그러나, 객체 인식은 사람만이 느끼는 펀크툼을 느끼지 못한다는 데에서 그 한계가 있다 (이재현, 2019). 결국 표면 상에서 드러나는 양식화되고 일반적인 코드 만을 읽어내고, 사용자가 이에 맞춰 사진을 읽어낼 때까지 자신의 방식을 훈련시키게 되는 것이다.

2. 스마트폰 사진

과거 디지털 카메라의 등장에 비해, 카메라가 부착된 스마트폰의 보급은 사진 촬영의 행동이 일상적으로 일어나는데 기여했다. 이에 따라 과거 인물 위주의 사진에서 나아가 가정에서 찍는 개인 사진의 대상도 생일이나 이벤트, 휴가 등으로 다양해졌다 (Broekhuijsen, 2017). 변화는 스마트폰 사진을 통해 촬영하는 대상과 그 활용 방식의 변화로까지 이어졌다. 특히 스마트폰 사진은 단순히 특별한 날 혹은 순간을 남기는 데서 나아가, 일상 생활에서의 기록들을 담아두는 역할을 도맡게 되었다. 카인드버그(Kindberg, 2005)는 스마트폰의 등장 후 나타난 사진첩의 변화에 대해 주목하고, 34 명의 사용자를 대상으로 사진첩 내부의 이미지와 촬영된 대상 대해 조사를 진행했다. 그 결과로 자신, 가족 혹은 지인, 혹은 타인의 사진으로 구성된 ‘사람이 포함된 사진’과 특정한 사물이나 실내 혹은 실외 경관, 애완동물, 그리고 스크린 혹은 필기가 드러나는 ‘사람이 아닌 다른 대상이 촬영된 사진’으로 구분하였다. 이후, 박병선은 스마트폰 사진의 또

다른 활용 가능성에 주목하고, ‘텍스트를 대상으로 촬영된 사진들’의 동기와 그 활용, 그리고 그 내용을 분석함으로써 스마트폰 사진의 활용의 변화에 대한 논의를 진행했다 (박병선, 2012). 더불어 SNS(Social Networking Service)의 등장 후, 자신이 촬영한 사진을 친구에게 공유하거나 공개된 SNS 에 포스팅하고, 공유된 사진을 자신의 스마트폰 사진첩에 저장하는 행태도 확인되고 있다 (Malik et al., 2016). 이러한 흐름에서 보았을 때, 현재 사용자의 스마트폰 사진첩의 내부는 직접 확인하지 않더라도 그 대상이 매우 다양하고 혼란스러운 상황임을 짐작할 수 있다.

3. 스마트폰 사진 관리와 찾기(Retrieval)

사진 관리 시스템에서 사진을 관리하는 방식은 곧 찾기의 방식과도 연결된다 (Whittaker et al., 2010). 스마트폰에서 사진을 찾는 가장 이상적인 방법은 사진에 있는 사람이나 촬영된 장소를 방문한 시점을 통해 사진을 찾는 방법일 것이다. 그러나 실내, 실외, 풍경, 사람, 물건, 이벤트 등 다양한 사진의 유형으로 인해 사진에 드러난 내용의 다양성은 무궁 무진 해진다. 이를 해결하기 위해 추가로 사용되는 데이터가 시간과 장소, 날짜, 콘텐츠에 직접 다는 태그와 같은 메타데이터이다 (Zhang et al., 2012). 그러나, 많은 사람들이 사진 앨범을 만들어 관리하는 일과 사진마다 직접 키워드 혹은 캡션을 다는 일은 수고스럽고 귀찮게 여기는 이유로, 사진 관리는 점차 소홀해지게 된다 (Sarvas et al., 2011).

이에 더해 사진 촬영의 대상이 다양해짐에 따라 사진을 되찾는 행위는 점차 결국 수많은 스마트폰 앨범의 사진들은 단순히 누적에 그쳤다. 그러나 최근 기술의 발전은 사진에 대한 태그나 주석을 형성하는 방식을 과거에 비해 빠르고 편리하게 만들었고(Van House, 2011), 현재에 이르러서는 기기가 자동적으로 이를 형성한 후 알아서 사진을 분류하고 정리하고 있는 것이다. 대표적으로는 시간과 장소와 같이 사진이 찍힌 맥락에 초점을 맞추는 방식과 사람이나 풍경, 대상과 같이 사진의 콘텐츠에 초점을 맞추는 방식이 있다 (Oliveira et al., 2018). 나아가 최근 인공지능 기술을 활용하여

이미지 주석을 자동으로 제공하는 기능에 대한 연구가 활발히 진행되고 있다. 대부분의 방법들은 각각의 사진에서 드러나는 콘텐츠를 분석하여 의미적 괴리를 줄이고, 이를 통해 검색 엔진에서도 사용될 수 있도록 자동 이미지 주석의 성능을 향상시키는 방법이다.

스마트폰 사진첩 내에서 개인 사진을 찾을 때 나타나는 질의 형태와 인터랙션 방식은 앞에서 살펴본 내용 중심의 검색 (CBIR)과 차이를 보인다. 먼저 일반적인 웹 검색에 비해, 개인 미디어에 대해 사용자는 무엇을, 누가, 언제, 어디서와 같은 정보에 더 관심을 보인다 (Jiang, 2017). 나아가 개인 사진 검색은 찾고자 하는 이미지에 대해 일반 웹 이미지 검색에 비해 명확한 정보 추구 형태를 띤다. 이러한 특성으로 인해 개인 사진을 찾을 때에는 시스템과의 인터랙션 시간에서 더 짧은 시간과 적은 인터랙션 횟수가 나타난다. 현재 사진 검색이 가지는 어려움을 미뤄보았을 때, 스마트폰 사진첩 내에서 일어나는 사진의 검색을 위해 개인 사진을 검색하는 행태와 그 학습 과정을 살펴보고, 이를 활용한 짧고 효율적인 검색 인터랙션의 설계가 필요함을 알 수 있다.

현재 사용되고 있는 사진 관리 시스템 중에서는 구글 포토(Google Photos)가 대표적으로 그 예가 될 수 있다. 구글 포토는 자사의 기술을 활용하여 사용자들에게 앨범을 자동으로 생성하고, 쓸모 없는 사진을 정리하여 제거 여부를 확인하거나 연속 촬영된 사진은 움직이는 사진으로 만드는 등 적극적으로 사진과 사진첩을 활용하도록 그 방안을 설계 중에 있다. 이 흐름에 맞춰 학계에서도 구글 포토의 새로운 분류 방식에 대해 관심을 갖기 시작했는데, 스마트폰 사진첩의 사진을 자동으로 정리하고, 쉽게 찾을 수 있도록 구글 포토가 제시하는 자동 사진 정렬과 분류, 그리고 태그를 다는 방법과 주석을 다는 새로운 방식에 대한 사례 연구가 등장하였다 (Nieuwenhuysen, 2018).

제 3 장 연구 문제

제 1 절 연구 문제의 설정

본 연구는 머신 러닝을 활용한 사진 찾기(Photo Retrieval)이 적용된 사진 관리 어플리케이션 내에서 시스템을 사용하는 사람들이 기기의 특성을 담은 전략이 제공되었을 때, 시간의 흐름에 따라 학습이 일어나는지 그 여부를 관찰하는데 목적이 있다. 나아가 제공된 전략은 얼마나 유효하게 사용되는지를 알아볼 것이다. 따라서 앞 장에서 살펴보았듯 객체 인식의 기본이 되는 컴퓨터 비전의 특성, 전략 형성을 위한 기기와 사람의 간극에 대한 이해, 사진의 특성을 살펴봄으로써 사진 검색의 경험에서 나타나는 학습의 형태와 전략의 활용을 통해 효율적인 검색을 위한 개선안을 모색해보고자 한다. 이에 학습의 과정을 검색에 소요된 시간과 시도 횟수, 태스크의 성공 여부, 사용 검색어를 통해 확인한 후, 제공된 전략의 수용 형태를 살펴보기로 하고, 다음과 같은 세 가지의 연구 문제를 설정하였다.

연구문제 1. 객체 인식 기술을 활용한 사진 찾기에서 학습의 양상이 나타나는가?

〈연구문제 1〉에서는 사용자의 관점에서 객체 인식을 활용한 검색 기능이 적용된 사진 관리 어플리케이션, 구글 포토(Google Photos)를 활용하는데 사진을 검색하는 데 있어 그 학습 과정을 관찰하고자 한다. 시도 횟수가 쌓이고, 시간이 지남에 따라 사용자가 구글 포토의 검색 기능을 어떻게 학습해가고 적응해 나가는지, 시간이 지남에 따라 검색 소요 시간에 변화와 검색 시도 횟수, 실패 횟수에 변화가 나타나는지 알아본다. 이를 위해 일주일간의 실험을 설계하고, 참가자들에게 사진 찾기 태스크를 매일 제공하여 태스크의 시작 시간과 종료 시간을 수집하여 측정된 두 시간의 차이를 도출할 것이다. 〈연구문제 1〉의 세부적인 연구문제는 다음과 같다.

연구문제 1.1

태스크의 수행 횟수가 누적됨에 따라 검색 소요 시간은 짧아지는가?

연구문제 1.2

태스크의 수행 횟수가 누적됨에 따라 검색 실패 확률은 줄어드는가?

연구문제 1.3

태스크의 수행 횟수가 누적됨에 따라 검색 시도 횟수는 줄어드는가?

이를 통해 사용자들이 머신러닝이 접목된 사진 찾기 행동에 학습 혹은 적응하는 모습을 보이는지, 학습이 잘 안되는 경우는 어떠한 경우인지 유추해볼 수 있을 것이다. 해석을 위해 각 태스크에 사용된 검색어의 변형을 추가적으로 수집한다. 더불어 단순히 태스크의 누적에 따른 변화 양상에서 나아가 일주일의 시간 축에서 변화의 추이를 살펴볼 것이다.

연구문제 2. 사용자의 학습 행태의 유형과 그 특성은 어떻게 나타나는가?

〈연구문제 2〉에서는 〈연구문제 1〉의 전반적 학습 행태 탐색에서 나아가, 개별 사용자의 사용 행태와 그 특성에 따라 유형을 분류하고자 한다. 시도 횟수의 누적에 따라 학습의 형태가 어떻게 달라지는지, 어떠한 이유에서 그 차이가 발생하는지 사용자에게 제공된 사진의 유형에 따라 그 차이를 살펴볼 것이다. 나아가 사후 설문을 통해 얻은 검색 경험의 어려운 점을 활용하여 학습 행태에 차이가 나타나는 이유에 대해 유추해본다.

연구문제 3. 사용자에게 주어진 검색 활용 전략은 받아들여지는가?

〈연구문제 3〉에서는 〈연구문제 1〉과 〈연구문제 2〉에서 살펴본 사용자의 관점에서 나아가 시스템의 관점에서 제공되는 검색 사용 전략을 참여자가 얼마나 활용하는지, 검색 경험의 누적을 통해 참여자가 직접 만들어내는

자신의 전략은 어떤 것이 있는지 알아본다. 나아가 추후 생겨난 사용자의 개별적인 전략의 사용 이유를 검색어의 활용과 추후 수집된 설문을 통해 분석한다. 이는 실험 기간 동안 챗봇을 통해 참여자가 응답한 데이터와 기재된 검색어의 사용, 그리고 사후 설문으로 수집한 자신만의 전략을 통해 확인해본다. 이에 따른 <연구문제 3>의 세부적인 연구문제는 다음과 같다.

연구문제 3.1

제공되는 팁은 태스크의 수행 중 얼마나 활용되는가?

연구문제 3.2

태스크의 시도 횟수가 누적됨에 따라 사용자는 자신만의 전략을 형성하는가?

연구문제 3.3

새로 만들어지는 전략은 어떤 이유로 사용되는가?

제 2 절 측정을 위한 개념 정의

본 절에서는 연구 문제에 대한 결과 분석을 위해 사전에 정의되어야 할 개념들에 대해 정의를 진행한다.

1. 객체 인식 기능을 활용한 사진 검색

본 연구의 연구 대상이 되는 ‘객체 인식 기능을 활용한 사진 검색’ 활동은 기존의 사진첩에서 활용하던 검색 기능이나 인스타그램 혹은 구글 이미지 검색과는 차이점을 가진다. 먼저 과거의 사진 관리 서비스에서는 사용자가 직접 태그나 주석을 달거나, 앨범을 제작하여 사진을 능동적으로 분류하는 행태를 보였다. 그러나 객체 인식 기능을 활용하는 사진 관리 서비스의 경우, 사용자가 업로드한 사진첩의 이미지를 기존에 학습된

모델을 바탕으로 읽어낸 후, 시스템 내부에서 사진을 분리하는 작업을 진행한다 (Nieuwenhuysen, 2018). 이에 따라 구글 포토에서의 검색은 시스템이 읽어낸 이미지 내부의 의미와 연관된 단어를 활용하게 된다. 즉, 사용자는 시스템의 내부에서 어떤 기준으로 해당 분류가 이뤄지고 있는지에 대한 정보를 얻기 어려운 상황에 놓이게 된다.

웹에서 일어나는 이미지 검색의 경우, 찾는 대상이 되는 이미지의 생성이 주로 타인이라는 지점에서 가장 큰 차이를 보인다. 특히 개인의 사진을 찾을 때에는 웹 검색에서보다 보다 시각적인 질의어를 더 많이 사용하는 모습을 보이며, 과거의 사진을 통해 ‘누가, 언제, 무엇을, 어디서’와 같은 정보를 추구하는 행태를 보인다 (Jiang et al., 2017). 이에 본 연구에서는 ‘객체 인식을 활용한 사진 검색’을 사용자의 스마트폰 사진첩 내에서 사진 관리 시스템이 사용자의 의도와는 별개로 분류한 단위를 통해 사용자가 맥락적, 시각적 검색어를 통해 자신의 사진을 검색하는 행위로 정의하겠다.

2. 학습

본 연구에서 해결하고자 하는 문제점은 현재 객체 인식 기술이 적용된 사진 검색에서의 경험이 불만족스럽고, 사용하기 어렵다는 데에서 기인한다. 이는 사용자가 객체 인식의 특성을 이해하지 못하고, 시스템은 현재의 한계에 대해 그 이유를 제공하지 않는 데에서 발생하는 것으로 유추해볼 수 있다. 초기 경험이 부정적이고, 사용에 어려움을 느끼게 되는 경우에는 시스템을 효율적으로 활용할 수 있는 방식을 학습하지 못하는 결과로 이어지게 된다. 이에 본 연구에서는 문헌 연구와 사전 조사를 통해 검색이 잘 활용될 수 있는 전략을 형성하고, 이를 사용자에게 제공하여 사진 검색의 방식을 학습하는지 그 여부를 파악하고자 한다. 나아가 궁극적으로는 사진 검색에서 나타나는 어려움을 파악하고, 거시적인 관점에서 사용 편의성(Usability)을 높이는 데 그 목적을 두고 있다.

본문에서 관찰하고자 하는 학습은 닐슨(Nielsen, 1994)의 정의에 따른다. 그에 따르면, 학습은 사용 편의성 속성 의 일부 중 하나로, 이를 학습성(Learnability)으로 표현하고 있다. 여기에는 사용자가 과업을 빠른 시간 내에 숙달할 수 있도록 시스템을 쉽게 배울 수 있게 만들어야 한다는 속성이 포함된다. 사람은 어떤 과업에 대해 반복할 수록 해당 과업을 수행하는 능력이 증대하는 모습을 보인다 (정광태 & 홍자인, 2008). 따라서 한 가지 과업의 수행 반복 횟수가 늘어남에 따라 과업 수행에 소요되는 시간이 줄어들을 유추해볼 수 있다. 단순한 시스템의 경우에는 단 몇 번의 수행 반복을 통해서도 쉽게 학습의 성과가 나타날 수 있지만, 본 연구에서 초점을 맞추고자 하는 객체 인식 기술이 접목된 사진 검색과 같이 그 내부를 알 수 없는 복잡한 시스템에서는 오랜 시간 여러 번 반복을 진행해야 학습의 효과가 나타나거나 혹은 학습이 일어나지 않을 수도 있는 것이다. 결국, 학습은 어떤 과업을 반복하면 나타나게 되는 시스템에 대한 이해와 숙달을 통해 그 과업 소요 시간이 점점 짧아지는 것을 뜻하게 된다.

제 4 장 연구 방법

본 연구는 실험 연구 방법을 사용한다. 실험 참가자는 일주일간 주어지는 태스크에 따라 사진을 찾도록 하는 실험을 수행한다. 실험을 통해 수집된 과업 수행에 소요된 시간 과 수집된 응답 데이터를 분석하였으며, 소요 시간의 변화에 대한 이해를 돕기 위해 사용된 검색어를 수집하여 추가 분석을 실시한다. 이후 사후 설문에 대한 답변을 통해 실험의 결과를 해석하는데 이를 활용한다. 제 1 절에서는 연구에서 활용한 연구 방법에 대해 소개한 후, 제 2 절에서는 실험 설계의 설계와 데이터의 수집에 대한 내용을 기재하였다.

제 1 절 연구 방법

1. 실험 연구 방법

본 연구의 주된 데이터 수집 방법은 실험 연구 방법이다. 실험 조사 방법은 선정된 소수의 인원을 바탕으로 통제된 환경 속에서 실험을 진행하여 분석 가능한 데이터를 비교적 손쉽게 얻을 수 있다는 장점을 가진 방법이다. 실험실 환경에서 높은 수준의 변수 통제를 바탕으로 데이터를 모을 수 있다는 것이 장점이지만, 반면 강도 높은 통제는 사용자의 자연스러운 행태를 관찰하는데 부정적인 영향을 미치게 된다. 나아가 일정한 기간을 두고 그 변화 양상을 확인하는 데에는 실험의 참여자가 매일 실험실에 방문하거나 비슷한 환경을 조성해야 한다는 데 그 어려움이 있다. 본 연구에서는 비교적 장기적인 시간의 축에서 일주일 간의 검색 태스크의 수행 과정을 살펴봄으로써 그 학습의 과정과 변화 양상을 파악해야 하기 때문에, 실험실 세팅이 가지는 단점을 보완할 필요가 있었다. 따라서 챗봇을 활용한 실험물 제작을 통해 실험 진행자와 참여자가 물리적으로 떨어져 있어도 일정 수준의 통제와 실험 진행이 가능한 방법을 고민하였다.

제 2 절 실험 방법

1. 실험 환경 설정

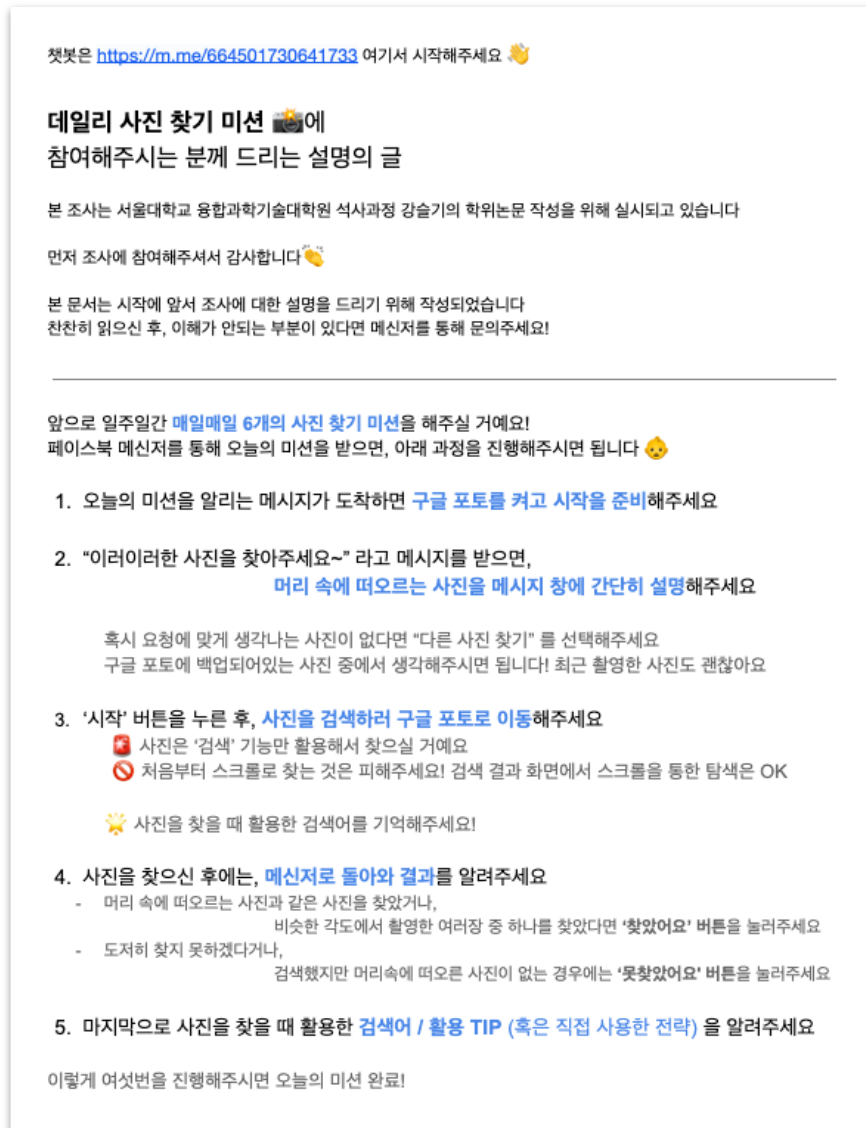
본 연구에서 실험은 진행자와 참여자가 물리적으로 떨어진 공간에서 실험 도구인 챗봇을 통해 일주일간 진행되었다. 원격으로 진행한 이유는 실험 진행자가 참가자의 개인 사진첩에 대한 접근하지 않고, 일주일 동안 진행되는 실험을 최대한 효율적으로 진행하고자 했기 때문이다. 이에 따라서 참여자가 실험의 과정을 쉽게 이해할 수 있도록 설명을 제공하는 가이드 문서와 실험 도구가 설계에 있어 주요하게 고려되었다.

1.1. 실험 가이드 준비

(1) 가이드 문서 제작

본 실험은 일주일간의 실험 조사 기간 동안 연구 진행자와 참가자가 대면하지 않고, 원격으로 이뤄진다. 따라서 참가자가 실험 진행 과정에 대한 이해를 도울 수 있는 명확한 가이드 문서가 필요하다. 실험 기간이 길고, 작성해야 할 항목이 여러 가지인 관계로, 모든 참가자가 이해하기에 쉽고 평이한 문장으로 작성되었다. 가이드 문서는 먼저 연구자에 대한 소개로 시작하여, 연구에 대한 간략한 소개를 제공한 후 사진 찾기 태스크의 진행 방법에 대한 설명을 제시한다. 다음으로는 구글 포토의 검색 기능을 활용하여 사진을 잘 찾을 수 있는 전략으로 팁(Tip)을 제공하고, 실험의 통제를 위해 주의하여 진행해야 할 사항들을 명시 해둔다. 이 때, 사진을 찾는 방법으로 스크롤을 통한 둘러보기 방식이 아닌 검색어를 통한 ‘검색’ 기능을 활용하여 사진을 되찾을 수 있도록 공지하였다. 마지막으로는 기본적으로 실험 진행에 대해 가질 수 있는 문의 사항과 프라이버시 및 개인 정보의 노출 가능성에 대한 설명을 질의 응답 형태로 작성하였다. 사진 찾기 태스크를 수행한다는 용어에서 참가자가 과업 수행에 대한 부담을 가질 수 있다. 이는 가이드 상에서 사진을 찾는 ‘미션’을 매일

수행하는 것으로 용어를 수정하여 과업 수행에 대한 참가자의 부담을 줄이고자 하였다. 제공된 가이드라인의 형태는 [그림 3]과 같았다.



[그림 3] 가이드 문서의 예시

(2) 팁(Tip)의 제작

현재 구글 포토에서는 검색 기능에 대해 짧은 설명 제공한다. ‘애완 동물(A pet)’, ‘가장 친한 친구(Your best friend)’, ‘좋아하는 도시(Your

favorite city)', 그리고 '지난 여름 참석한 결혼식 (A wedding you attended last summer)' ② 와 같이 사용자가 검색할 수 있는 대상을 개략적으로 제시하는 데 그치고 있으며, 그 이상의 검색 활용 방안에 대해서는 구체적으로 밝히고 있지 않다. 특히, 현재 알고리즘을 통한 머신 러닝을 활용한 서비스는 불가피한 오류의 경험과 그 블랙 박스(Black Box)와 같은 특성으로 인해 서비스를 접하는 사용자가 시스템에 대한 일정한 멘탈 모델을 형성하기 어렵다 (Zhou & Chen, 2018).

본 연구에서는 사용자가 머신 러닝의 특성을 이해하고, 사진 검색을 보다 쉽고 유용하게 활용하는 것을 목표로 하고 있으므로, 사용자가 객체 인식과 머신 러닝 내부의 특성을 먼저 인지하고, 나아가 학습할 수 있도록 검색의 전략, 즉 팁(Tip)을 작성하여 제공하였다. 작성을 위해 참고할 수 있는 기존의 가이드라인이 없는 상황이었기에, 기존의 문헌에서 밝힌 객체 인식 기술이 사람보다 오류를 잘 일으키는 경우(Russakovsky, 2015)를 참고한 후, 이에 더하여 직접 진행한 사전 조사를 통해 얻은 사용자가 '유용하게 생각한' 전략에 추가하여 총 여섯 가지의 팁을 작성하였다.

가장 먼저 '(,콤마)의 사용'은 사전 조사에서 가장 유용하게 활용된 전략으로, 참여자가 콤마의 활용이 가능함을 발견한 후 가장 유용하게 사용한 전략으로 기술한 것이다. 검색하고자 하는 대상을 지칭하는 명확한 단어를 사용하는 경우에 원하는 검색 결과를 얻을 가능성이 높다. 그러나 어떤 단어를 활용할 수 있을지 모르는 경우에는 검색 가능한 범위를 탐색하는 방법이 유용하게 사용될 수 있다. 이 때 활용될 수 있는 기능이 ,(콤마)를 활용하여 사용된 질의어가 모두 포함된 사진을 찾는 것이다. 단순히 여러 개의 사물을 콤마로 활용하여 그 범위를 좁히는 것 보다 사람과 사물, 시간과 사물과 같이 검색어를 혼합하여 사용하는 경우에 가장 유용하게 활용된다.

② 구글 포토 웹사이트의 'Help' 페이지 내 활용 검색어 예시

https://support.google.com/photos/answer/6128838?hl=en&ref_topic=6128818&co=GENIE.Platform%3DDesktop&oco=1

두 번째로 ‘남자’ 혹은 ‘여자’를 검색어로 활용’은 구글 포토가 객체 인식을 통해 사진에 드러난 얼굴을 인식하고, 비슷한 사람의 얼굴을 한데 묶었지만 이름을 따로 지정하지 않았을 경우 활용될 수 있다. 인물의 이름은 구글 포토가 학습하지 못하고, 많은 사람들이 사진첩 내에서 자동으로 분류된 인물에 대해 이름을 정하지 않는다는 지점에서 유용하게 활용될 수 있는 전략이다.

세 번째로, 검색어를 활용할 경우에는 ‘상위 개념의 단어’ 혹은 ‘굉장히 자세한 단위의 단어를 활용’하도록 한다. 이는 사람이 대상에 대해 이름 부르는 분류 체계와 머신 러닝이 피처를 통해 학습한 후 특정 피처를 가진 대상에 대해 라벨을 붙이는 단위에서 발생하는 차이에 대해 사람들로 하여금 기기의 방식을 학습하도록 하기 위해 제공된다.

네 번째로, ‘화면에서 볼 수 있는 것을 검색어 활용’에 관한 팁은 객체 인식이 화면을 통해 인식할 수 있는 범위에 대한 특성을 알리기 위해 작성되었다. 일례로 객체 뿐만 아니라 화면의 색상을 활용할 수 있는 데에서 그 활용 가능성을 알리기 위함이다. 그러나, 추상적인 분위기 같은 경우에는 검색이 불가능함을 공지한다.

다섯 번째로, ‘이벤트의 검색’은 객체 인식의 검색 가능 범위에 대해 사용자에게 인식시키고자 제공되는 전략이다. 실제로 사용자의 사진첩 내부의 사진을 구글 포토가 분류한 방식은 ‘사물(Things) 폴더’에서 모두 확인할 수 있다. 그러나, 실제로 사물 폴더에 들어가보면 단순히 객체 만을 분류하고 있지 않다. 미팅, 마라톤, 하이킹과 같이 화면에서 읽어낼 수 있는 ‘활동’과 ‘이벤트’도 읽어낼 수 있는 것이다. 사전 조사를 통해 구글 포토가 단순히 사물 만을 인식하는 줄 알았던 한 참여자는 활동의 분류가 사물 폴더에 있는 것에 대해 의아함을 표했다. 이렇듯 구글 포토가 인식 가능한 범위에 대한 인지를 위해 다섯번 째 전략을 제공한다.

마지막으로 ‘사물 - 분류의 탭에서 검색어 참고하기’는 구글 포토 내에서 자동으로 분류되어 알 수 없는 분류 체계를 간접적으로 확인하고, 검색어로 활용할 수 있음을 상기시키기 위한 전략으로 제공되었다 [그림 4].

👉 구글 포토를 활용한 사진 찾기 TIP 👉

- 1 , (콤마) 를 활용할 수 있어요
 ** 시간 - 사물, 장소 - 사물 과 같이 다른 형태의 검색어를
 같이 사용하면 더 잘 찾아집니다!
- 2 사람이름을 정해두지 않았을 때에는
 ‘남자, 여자’를 검색어로 활용하여 찾을 수 있어요
- 3 명확한 물체를 찾을 때에는,
 상위 개념의 단어를 검색어로 활용하거나
 굉장히 자세한 단어를 활용합니다
 ex. 정확하고 세부적인 길이름, 건물명 : 경리단길, 부채길, 쓸비지
- 4 사물 외에도,
 화면에서 ‘볼 수 있는 것’을 검색어로 활용해주세요
 ex. 빨간색, 회색과 같은 화면의 색상
 느낌적인 느낌, 형용사는 구글 포토가 찾지 못해요
- 5 ‘이벤트’도 검색이 가능해요
 영어로 쉽게 부를 수 있는 검색어를 활용해주세요
 ex. 미팅, 마라톤, 백패킹 등
- 6 분류 - 사물에서 검색 가능한 단어를 참고해주세요

하단 바의
'설정'에 들어가신 후



앨범

➡



사물



동영상



콜라주

[그림 4] 제공된 팁(Tip)의 예시

실험 진행 중에는 제공된 전략을 활용하도록 하였다. 검색의 경험이 누적될수록 기존의 팁에서 자신의 방식에 맞게 변형하거나 전략을 새로 만들 경우에는 챗봇에 공유하도록 요청하였다. 이후 태스크 수행 중에도 언제나 이 팁을 활용할 수 있도록 [그림 5]와 같이 챗봇의 메뉴 목록에

올려 두었다. 메뉴 중 ‘TIP 보러가기’를 선택하면 앞서 제공되었던 팁의 문서가 즉시 제공되었다.



[그림 5] 챗봇의 메뉴에서 확인할 수 있는 팁의 예시

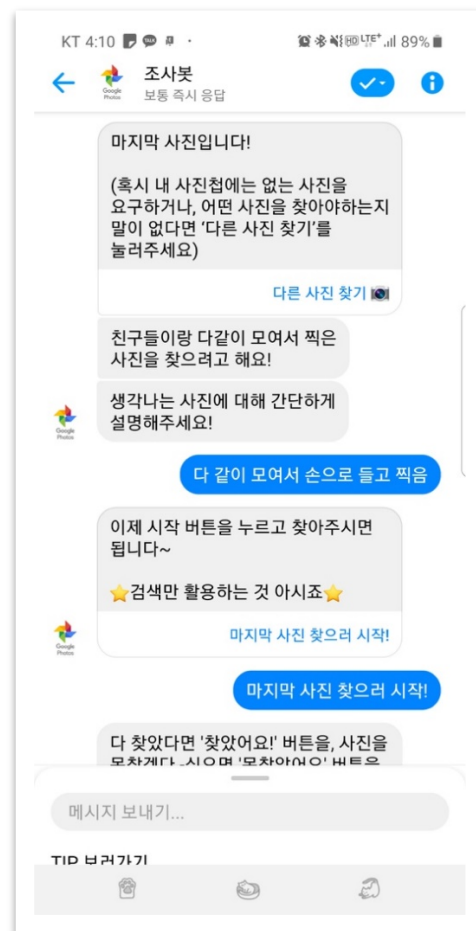
1.2. 실험 도구 제작

(1) 실험용 챗봇 제작

일주일 동안 매일 태스크를 수행해야하는 본 실험에서는 참여자의 꾸준한 참여가 매우 중요하다. 그러나, 조사 기간에 걸쳐 참가자가 매일 실험에 자발적으로 참여하기는 어려운 일이다. 실험 조사용 웹페이지를 제작하여 사용자들에게 제공할 수 있지만, 매일 직접 참여하기를 기대하고, 이를 확인하여 지속적인 참여를 요청하는 데에도 한계가 따른다. 이에 따라 장기간에 걸쳐 참여를 도모하면서도 원격으로 데이터를 수집하기에 적당한 조사 도구가 요구된다.

챗봇은 메시지를 알림 시간에 맞춰 참가자에게 전달할 수 있다는 장점을 갖고 있다. 그 외에도 원격으로 이뤄지는 실험에서도 진행자와 참가자 사이의 질의응답과 같은 의사소통이 활발하게 이뤄질 수 있다는 점에서 적합한 실험 진행 도구이다. 챗봇을 조사 도구로 활용하려는 시도는 최근의 연구에서 활발히 이뤄지고 있다. 특히 김수민(Kim et al., 2019)은 챗봇을 통한 데이터의 수집이 웹페이지를 통한 설문 조사의 데이터에 비해 더

다양하고 좋은 질의 데이터를 얻을 수 있으며, 참가자와 챗봇 사이의 인터랙션을 통해 기존의 서베이 방식보다 더 적극적으로 조사에 참여함을 밝혔다. 이에 본 실험에서는 챗봇을 활용하여 참가자에게 태스크 수행을 위한 알리를 제공하고, 이를 통해 실험 참가를 유도하여 데이터를 수집하는 방식을 택하였다. 이후 실험을 위해 챗봇 빌드 플랫폼 Manychat^③을 통해 페이스북 메신저 (Facebook Messenger)를 기반으로 하는 챗봇을 제작하였다 [그림 6].



[그림 6] 실험에 활용된 챗봇의 예시

챗봇을 통해 한 번의 태스크를 진행하는 순서는 다음과 같다.

^③ <https://manychat.com/>

- ① 먼저 n번째 일차를 시작하게 되면, 첫 번째 태스크로 참가자에게 이전 회차에 제공되지 않은 새로운 태스크가 제공된다.
- ② 태스크를 제공받은 후, 머리 속에 떠오른 사진에 대해 간단한 문장으로 설명을 남긴다. 이 때, 태스크에 알맞은 사진이 없거나 생각나는 사진이 없는 경우에는 ‘다른 사진 찾기’ 버튼을 눌러 새로운 태스크를 받는다.
- ③ ‘시작하기’ 버튼을 누른 후, 구글 포토로 이동하여 앞서 머리 속에 떠오른 사진을 찾게 된다.
- ④ 이후 생각한 사진과 같은 사진을 찾았을 경우에는 챗봇으로 돌아와 ‘찾았어요’ 버튼을 누르게 되고, 앞서 생각한 것과 똑같은 사진은 아니지만 같은 대상, 혹은 반복되는 시도 후 사진을 찾지 못하는 경우에는 ‘못 찾겠어요’ 버튼을 누른다.
- ⑤ 검색 시 검색을 시도한 횟수를 적는다.
- ⑥ 검색 시 활용한 검색어들을 대화창에 작성한다.
- ⑦ 제공된 팁 중 이번 태스크에서 활용한 팁을 선택하거나, 자신이 변형 혹은 새로 사용한 전략을 대화창에 적는다.
- ⑧ 다음 태스크가 제공된다.

한 번의 태스크를 여섯 번 반복한 후, 그날의 태스크를 진행하며 검색이 어려웠던 점 혹은 개선되어야 할 부분과 그 이유에 대한 짧은 질문을 제공한다. 이렇게 하여 하루에 수행되어야 할 태스크는 마무리된다.

(2) 사진 찾기 태스크 제작

참가자는 실험 참가 기간 동안 매일 챗봇을 통해 태스크를 제공받고, 자신의 스마트폰 사진첩 내에 있는 사진을 찾게 된다. 참가자에게 사진을 찾도록 요청할 때 사용되는 태스크를 선정하기 위해서는 먼저 태스크를 요청하는 방법과 두번째로 어떤 사진을 찾도록 할 것인가에 대한 고민이 선행되어야 한다. 첫 번째로 사진을 찾는데에는 두 가지 방법을 활용할 수

있다. 먼저 개인 사진첩에 조사자가 직접 접근하여, 사진첩 내의 사진 하나를 랜덤으로 선택하여 참가자에게 같은 사진을 찾도록 요청하는 인식 기반(Recognized-Based)의 방식과, 검색을 위해 과제를 제공하는 방식을 사용되는 태스크 기반(Task-Based)의 방식(Kofler & Lux, 2009)이다. 먼저 전자의 방식은 실험 진행자가 참가자의 개인 스마트폰 사진첩에 직접적으로 접근한다는 데에서 프라이버시 침해를 불러일으킬 수 있다. 실험 진행자가 직접적으로 참가자의 스마트폰 사진첩에 접근하는 경우를 최소화하고, 참가자가 민감하지 않도록 배려하는 방식에 대한 고려가 필요했다.

실제 네 명의 참가자를 대상으로 한 사전 조사를 통해 두 방식에 대한 참가자의 프라이버시 노출에 대한 우려의 목소리를 듣고, 적합한 사진 회고 방식을 파악할 수 있었다. 두 명의 참가자 에게는 사진첩 내 공개가 괜찮은 부분을 캡처한 후, 사생활이 노출될 가능성이 있는 사진을 가리는 조치를 취하여 진행자에게 전달하는 방식을 요청했고, 나머지 두 명에게는 사진을 찾는 상황을 가정하는 태스크를 작성하여 제공하였다. 전자의 참가자는 사진첩 내부를 캡처하는 단계에서부터 사진첩 공개에 대한 부담감을 느꼈음을 밝혔으며, 랜덤으로 진행자에게 전달할 사진을 선정할 선정하는 데 있어서도 고심하는 모습을 보였다. 더하여, 이후 사진을 검색하는 데 있어 목적이나 맥락이 없이 과거에 자신이 촬영한 사진을 찾는 데에만 그 목적을 두고 있어 실제의 검색 환경과 상이할 수 있다는 가능성을 확인하였다. 반면, “축제나 콘서트, 연극과 같은 이벤트에서 촬영한 사진이 있나요?”와 같은 태스크를 제공한 참가자는 실험 진행자에게 자신의 사진첩을 직접 노출할 필요가 없었으며, 검색 과정 또한 실제의 방식과 유사한 모습을 보였다.

어떤 사진을 검색하도록 태스크를 구성할 것인가에 대한 고민으로는, 먼저 카인드버그의 연구에서 도출된 스마트폰 사진의 대상에 따른 분류를 기준으로 활용하였다 (Kindberg et al., 2005). 이후 국내 리서치 회사

엠브레인 (EMBRAIN)의 스마트폰 사진의 촬영 목적 순위 ^④에 따라 가장 높은 순위를 참고하였다. 사진은 촬영자의 목적에 따라 촬영하게 되며, 사진첩에는 이렇게 목적에 따라 만들어진 사진이 누적된다. 따라서 스마트폰 촬영 목적에서 드러난 순위와 유사하게 스마트폰 사진첩에는 촬영된 대상의 사진이 많음을 유추할 수 있다.

프레젠테이션 작성을 위한 이미지 검색과 같이 웹에서 나타나는 검색 과업과는 그 목적과 검색 행태에서 차이를 가지게 된다 (Jiang, 2017). 특히 기존의 이미지 찾기(Image Retrieval) 연구에서 활용된 이미지 찾기 과업과는 달리, 본 연구에서는 스마트폰에 누적된 개인의 사진을 찾는다는 데 그 차이점이 있다. 따라서 사용자의 질의도 다르게 나타날 수 있다. 이와 같은 개인 사진 검색의 특성을 고려하여, 태스크를 제공할 때에는 일상에서 발생 가능한 상황으로 한정하여 최대한 자연스럽고 개인적인 질의를 이끌어낼 필요가 있었다. 태스크를 작성하기 위해 개인 사진을 검색하는 목적을 탐색하고자 실시한 사전 연구를 통해 수집된 개인 사진 검색의 목적과 수집 상황에 대한 서술을 참고하였다. 22 명의 참여자를 대상으로 일주일간 구글 포토(Google Photos)의 사용 기록과 설문을 수집한 데이터를 통해 개인 사진을 검색할 때 활용하는 검색어 및 그 사용 맥락을 파악할 수 있었다. 구글 포토를 활용한 사진 검색의 목적으로는 크게 ‘특정 사진을 취하기’, ‘둘러보기’, ‘정리하기’, 그리고 ‘시험하기’로 나뉠 수 있었다. 그 중에서도 특히 여행 시 필요한 여권 사진을 찾거나 친구에게 공유하기 위한 특정한 사진을 찾고자 하는 목적에서 활용된 검색 사용의 맥락을 활용할 수 있었다. 더하여 앞서 살펴본 스마트폰의 사진첩을 구성하고 있는 대상의 분류를 활용하였다. 이를 통해 [표 2]와 같이 태스크를 작성하였다.

사진의 개수는 일주일 동안 참가자가 매일 여섯 개의 사진을 찾을 수 있도록 총 42 개의 태스크를 제작하였다. 이후 태스크로 제공한 사진이

^④ 2017 디지털 카메라 vs 스마트폰 카메라 관련 인식 조사, EMBRAIN 트렌드모니터.(2017)

<https://www.trendmonitor.co.kr/tmweb/trend/allTrend/detail.do?bldx=1576&code=0102&trendType=CKOREA>

참가자의 사진첩에 없거나, 떠오르는 사진이 없을 경우 다른 사진을 찾을 수 있도록 21 개의 추가 태스크 세트를 제작하였다.

[표 2] 제공된 사진 찾기 태스크의 예시

대분류	소분류	예시
사람을 찍은 사진	한 명의 가족 / 친구	친구가 웃기게 나온 사진을 찾아 친구에게 카톡으로 공유하고 싶어요.
	여러 가족 / 친구	결혼식이나 친척들과 모였을 때 찍은 사진을 찾고 싶어요.
	사람 대상	축제나 콘서트, 연극과 같은 이벤트에서 촬영한 사진이 있나요?
	다른 사람이나 그룹	단체 사진을 공유받거나, 찍어둔 사진이 있나요? (앨범 사진도, 행사 기념 사진과 같은 단체 사진도 괜찮아요)
사람이 아닌 대상을 찍은 사진	특정한 대상	예쁘거나 맛있는 디저트도 사진으로 담아 두시나요? 맛있게 먹은 디저트의 사진을 찾아주세요.
	야외 사진	배경화면에 쓰면 좋겠다 싶은 풍경사진을 찾으려고 해요 (배경화면으로 지정 해두기 위해 다운받은 사진도 괜찮아요)
	스크린, 글귀 등에 대한 이미지	발표자료 혹은 회의자료에 필요한 회의 증빙 사진을 찍은 적이 있나요?
	애완동물	반려동물을 키우시나요? 웃겨서 찍어둔 우리집 반려동물의 사진을 찾으려고 해요. (반려동물이 없다면 저장한 동물 사진도 괜찮아요)
	실내	분위기 좋은 카페 혹은 공간을 사진으로 남겨둔 적이 있나요?
	기타	인스타그램에 올릴 평범한 일상 사진을 찾으려고 해요.

1.3. 참여자 모집 및 선정

본 연구는 시간에 따라 변화하는 사용자의 사진 검색의 학습 과정과 전략의 활용을 살펴보는데 그 목적을 두고 있다. 따라서 이에 적합한 조사 대상자는 구글 포토 내에서 검색 기능을 한번도 사용해보지 않은 사람으로 한정하는 것이 바람직하다. 설문을 통해 구글 포토를 사진 백업 및 관리 어플리케이션으로 활용하고 있지만, 검색 기능을 한번도 활용해보지 않은 참가자를 선정의 조건으로 제한하였다. 더불어 앞서 참고한 스마트폰 사진 촬영 현황 조사에서 나타난 바와 같이, 스마트폰 사진의 활용 빈도가 가장 활발한 20-30 대를 선정 기준으로 설정하였다.

이에 더해, 사전 조사를 통해 사진을 검색하는데 그 결과가 사진첩 내 사진의 개수에 영향을 받는다는 점을 확인하였다. 일례로 사진의 양이 너무 적은 경우, 제공되는 태스크에 부합하는 사진이 없을 경우가 발생할 수 있다. 이에 따라 사진첩 혹은 구글 포토의 보관함 안에 사진이 5000 개 이상인 사람을 선정하였다. 이를 바탕으로 제작한 사전 설문을 통해 모집한 총 인원은 23 명(여 15, 남 8) 이었으나, 일주일의 실험 진행 중 중도 포기를 하는 경우를 제외하여 최종적으로는 총 16 명(여 13, 남 3)의 데이터를 수집할 수 있었다 [표 3].

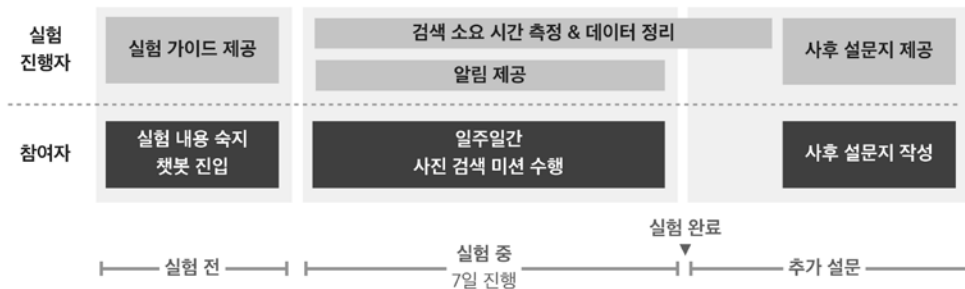
[표 3] 실험의 참여자 목록

PID	성별	나이	스마트폰 / 구글 포토 내 사진 개수
P01	여성	26 - 30	5001 - 10000
P02	여성	26 - 30	10001 - 25000
P05	여성	26 - 30	25001 - 50000
P06	남성	19 - 25	5001 - 10000
P07	여성	26 - 30	10001 - 25000
P10	여성	19 - 25	10001 - 25000
P11	여성	26 - 30	5001 - 10000

P12	여성	19 - 25	5001 - 10000
P13	여성	31 - 35	10001 - 25000
P14	여성	19 - 25	10001 - 25000
P15	여성	26 - 30	5001 - 10000
P17	남성	26 - 30	5001 - 10000
P18	여성	19 - 25	5001 - 10000
P20	여성	26 - 30	10001 - 25000
P21	남성	31 - 35	10001 - 25000
P23	여성	31 - 35	5001 - 10000

2. 데이터 수집

데이터 수집 과정에서는 제작된 챗봇으로 수집되는 태스크의 시작 시간과 완료 시간의 타임 스탬프, 태스크에 따라 검색 대상이 되는 사진의 설명, 검색 시도 횟수, 사용된 검색어, 그리고 검색 시 사용된 전략의 수집이 이루어졌다. 실험의 목적은 태스크의 시행 횟수에 따라 검색 수행 시간의 변화를 관찰하고, 검색 시 사용되는 전략에 대해 사용의 이유와 전략의 발전 과정을 살펴보는 것이다. [그림 7]은 실험의 진행 과정 중에서 나타나는 진행자와 참여자의 역할, 그리고 그에 따른 실험의 프로세스를 도식화한 것이다.



[그림 7] 진행 과정에 따른 실험 진행자와 참여자의 역할

2.1. 실험 과정

(1) 실험 전

실험 참여자는 전달받은 구글 닥스(Google Docs)의 링크를 통해 실험의 가이드라인을 숙지한 이후, 실험이 진행되는 챗봇으로 진입할 수 있다. 챗봇에 들어오게 되면, 사용자는 별도의 유저 아이디를 할당 받게 된다. 이를 통해 참가자 개인을 구분하되, 직접적인 개정의 노출을 우회할 수 있었다. 본격적인 조사 시작 전, 튜토리얼 형식의 검색 태스크를 제공하여 머신 러닝이 활용된 사진 기능을 한번도 사용해보지 않은 사람들로 하여금 검색의 기능을 익힐 수 있도록 하였다. 튜토리얼의 목적은 검색 기능이 익숙지 않은 사람들에게 사진 검색의 기본적인 기능을 숙지 시키기 위함이었다. 나아가 사진 검색의 경험을 모두 동일선상에 두기 위한 장치로 활용하기 위해 간단한 튜토리얼을 먼저 진행하도록 요청하였다.

(2) 실험 중 - 일주일 동안의 태스크 수행 중

참여자는 일주일간 매일 자신이 편한 시간에 챗봇의 대화방에 들어와 여섯 개의 사진 찾기 태스크를 완수한다. 챗봇을 활용하여 원격으로 조사가 진행될 때 통제가 가장 어려웠던 점은 참가자의 지속적인 참여를 유지하는 일이었다. 따라서 참여자가 꾸준히 실험에 참여할 수 있도록 알림이 필요했다. 이는 매일 자동적으로 제공되는 챗봇의 알림과 더불어, 그 날의 태스크를 수행하지 않은 참여자를 확인한 후 실험 진행자가 직접 알림을 제공하는 방식으로 해결할 수 있었다.

(3) 실험 후

일주일간의 사진 찾기 과업을 마친 후에는, 참여자는 실험 진행자가 제공하는 간단한 설문지에 답하게 된다. 먼저 일주일간의 전반적인 사진 검색 경험에 대한 만족도를 5 점 척도로 수집한다. 그리고 해당 점수에 대한

이유를 수집하게 된다. 다음으로 참여자는 사진을 검색하면서 느꼈던 가장 어려운 점을 기술한 후, 구글 포토에서 검색을 지속하며 만든 자신만의 새로운 전략을 작성하게 된다. 추가적으로 사진 검색 태스크의 맥락이 이해되지 않는 경우에는 별도의 질문을 형성하여 추후 데이터의 결과를 해석하는데 활용한다.

2.2. 데이터 수집과 분석

최종 분석 대상이 된 데이터는 일주일간 매일 6 개의 사진 찾기 미션을 모두 수행한 총 16 명(여 13, 남 3)의 실험 참여자로부터 수집된 결과이다. 본래 사전 설문을 통해 23 명이 실험 참여 대상으로 선정되었으나, 실험 완료 기간 중 중도 포기한 7 명 (여 2, 남 5)의 데이터를 제외한 후 수집된 데이터에 대해 분석을 진행하였다.

가장 주요한 분석 대상은 검색 시작 시간과 완료 시간 사이의 검색 소요 시간이다. 이는 검색 완료 시간에서 검색 시작 시간을 제한 후, 그 값으로 도출된 소요 시간을 초 단위까지 수집하였다. 이렇게 수집되는 소요시간은 각 참여자마다 첫번째 태스크에서부터 마지막 42 번째 태스크까지 누적되어 추후 분석에 활용되었다. 검색 성공 여부의 경우, 참여자로 하여금 머리 속에 떠오른 사진을 찾았을 경우 챗봇을 통해 ‘생각한 사진을 찾았어요’버튼을 누르도록 하였으며, 사진을 찾지 못할 경우에는 ‘못 찾겠어요 버튼을 누르도록 요청하였다. 이후 분석의 용이를 위해 성공의 경우를 ‘1’로, 검색 실패의 경우는 ‘0’으로 데이터를 수집하였다. 입력된 검색어는 추후 분석에서 학습의 변화 과정을 내용적으로 분석할 때 사용하기 위해 수집되었다. 다음으로는 주어진 전략 중 참여자가 사진을 검색할 때 활용한 전략을 선택하거나, 자신이 직접 경험한 후 만들어낸 전략을 작성하여 별도로 수집될 수 있도록 하였다. 이렇게 수집된 데이터는 [그림 8]과 같다. 나아가 매일 여섯 개의 태스크가 완료되면, 그날의 검색에서 어려웠던 부분에 대해 작성을 요청한다. 이는 텍스트의 형태로

추후 데이터를 해석할 때 검색을 학습하는 과정에서 겪은 어려움에 대해 이해를 돕고자 수집되었다.

태스크 수행 횟수의 누적에 따른 검색 소요 시간의 변화 및 검색 성공률, 소요 시간의 유의미한 상관 관계를 살펴보기 위해 각 변수에 대해 단순 회귀 분석을 실시하였으며, 유효성 검사를 통해 결정계수의 유의미함을 확인할 수 있었다.

trial #	PID	task	사진설명	성공여부	시도 횟수	검색어	TIP	duration
3	P18	2729711480379< 데이트 코스로 추천 롯데월드		1	4	놀이공원,사람, 회전목마, 커풀	가장 부피가 큰물체위주, 배경은 적어도 5	0:00:16
3	P20	2163433713706< 친구가 엄청 잘 나온 노가리 들고있는 현		0	7	노가리, 아외, 테라스, 맥주, 테이블	화면에 드러나는 검색어 활용하기	0:01:08
3	P21	1882985078469< 여행지에서 먹은 가 클래식고에서 먹은		1	1	클래식고, 음식	, (곰마) 사용하기	0:00:15
3	P23	2921257851225< 친구들과 만나 재밌 오키나와 코우리 대		1	1	코우리, 일본, 바다	, (곰마) 사용하기	0:00:29
3	P12	2473190296058< 전에 저장해둔 유용 피부 좋아지는 방법		0	3	피부, 뷰티, 동안	정확한 명칭 사용하기	0:00:27
4	P01	2633262123368< 엄마랑 찍은 사진을 엄마랑 여수나 전주		1	1	엄마	인물(여자,남자) 사용하기	0:00:21

[그림 8] 수집된 데이터의 예시

사후 설문은 실험 기간동안 느낀 사진 검색의 전반적인 어려움과 자신이 형성한 전략에 대한 이유를 얻기 위해 짧은 질문 형식으로 요청되었다. 질문 내용으로는 검색의 경험 중 가장 어려웠던 점은 무엇이었는지 설명을 요청했다. 이후 검색이 가장 유용하게 활용될 수 있는 경우에 대해 질문하였다. 이는 가장 쉽고 효율적으로 검색을 활용했다고 느낀 경험을 우회적으로 수집하고자 함이었다. 나아가 시간이 지나면서 사진을 검색하는 활동이 점점 쉽게 느껴졌는지 참여자의 주관적인 느낌에 대해 서술을 요청하였다. 마지막으로 조사를 마무리하며 자신이 새롭게 만들어낸 검색 효율을 높일 수 있는 검색 전략에 대해 추천을 요청하였다. 이렇게 수집된 설문의 응답은 실험 결과 데이터에서 드러나지 않은 전략 사용 및 형성의 이유와 구글 포토내에서 사진을 검색할 때 나타나는 어려움에 대한 세부적인 이유를 분석할 때 활용되었다.

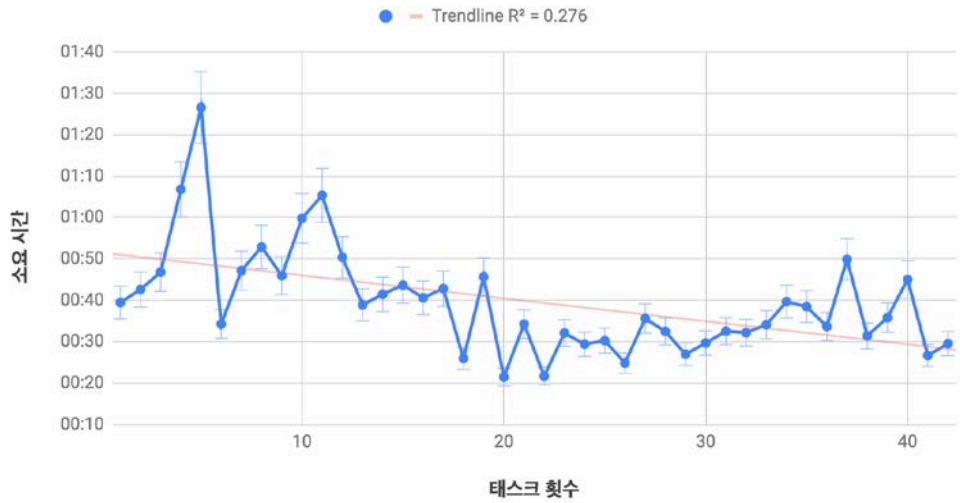
제 5 장 연구 결과

제 1 절 객체 인식을 활용한 사진 찾기의 학습

본 절에서는 〈연구문제 1〉 객체 인식 기술을 활용한 사진 찾기 시스템에서 학습의 양상이 나타나는지 그 여부를 확인하고자 한다. 실험을 통해 일주일간 수집된 태스크의 소요 시간이 태스크 시행의 누적에 따라 나타나는 변화의 양상을 살펴보았다.

1. 시도 횟수의 누적에 따른 소요 시간의 변화

본 연구에서는 머신러닝이 접목된 사진 찾기 시스템에서 발생할 수 있는 외부적인 요인에도 불구하고, 반복적인 사용을 통한 ‘학습’의 형태가 나타나는지 확인하기 위해 42 회의 태스크를 매일 6 개씩 배분하여 일주일 간 제공 하였다. 〈연구문제 1.1〉의 결과를 알아보기 위하여 매일 16 명의 참가자가 일주일 동안 진행한 태스크로부터 총 672 개의 소요 시간 데이터를 수집할 수 있었다. [그림 9]는 이를 태스크의 시행 횟수 축에 나열한 것이다. 수집된 데이터를 바탕으로 전체 참가자의 평균 소요 시간의 추이를 보았을 때, 태스크의 실행 횟수가 누적됨에 따라 평균 소요 시간이 줄어드는 모습을 확인할 수 있다. 시간의 누적에 따른 평균 소요 시간이 유효하게 영향을 미침을 확인하기 위해 단순 회귀 분석을 실시한 결과, 태스크 횟수의 누적에 따라 평균 소요 시간은 유의미하게 줄어듦을 확인할 수 있었다 ($R^2=0.276$, $p=0.000$).



[그림 9] 태스크 누적 횟수에 따른 소요 시간의 변화

그러나, 세부적인 태스크 수행 횟수에 따른 소요 시간과 표준 편차를 나타낸 [표 4]에서는 각 태스크마다 전반적으로 분산이 큰 분포를 보이고 있다. 이렇듯 불규칙적인 형태와 각 값들이 넓은 분포를 보이는 형태는 이후 <연구문제 2>에서 개별 사용자의 학습 행태에서도 두드러지게 나타나는데, 사후 설문을 통해 검색의 대상이 되는 사진에 따라 소요 시간이 상이해질 수 있으며, 이에 따라 머신 러닝이 제공하는 검색의 결과를 예측하기 어렵다는 점을 그 이유로 유추해볼 수 있다.

[표 4] 태스크 수행 횟수에 따른 평균 소요 시간과 표준 편차

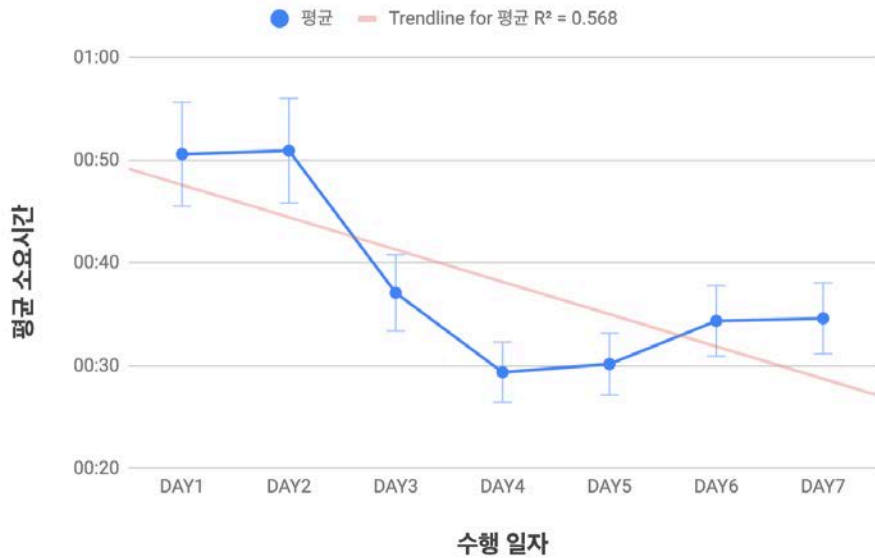
수행#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
평균	00:39	00:43	00:47	01:07	01:27	00:34	00:47	00:53	00:46	01:00	01:05	00:50	00:39	00:41	00:44	00:41	00:43	00:26	00:46	00:21	00:34
SD	00:41	00:30	00:51	00:51	02:13	00:21	00:41	00:42	00:28	01:21	01:23	01:06	00:29	00:46	00:32	00:33	00:36	00:23	00:40	00:15	00:32
수행#	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
평균	00:22	00:32	00:29	00:30	00:25	00:36	00:32	00:27	00:30	00:33	00:32	00:34	00:40	00:38	00:34	00:50	00:31	00:36	00:45	00:27	00:30
SD	00:17	00:29	00:30	00:22	00:17	00:34	00:23	00:24	00:18	00:25	00:23	00:32	00:26	00:47	00:27	00:42	00:22	00:26	00:28	00:29	00:26

본 실험은 하루에 6 번의 태스크를 일주일 동안 수행하는 방식으로 진행되었다. 따라서 태스크 수행 횟수의 누적을 거시적인 시간의 축에서 바라보았을 때, 수행 일자가 누적됨에 따라 학습이 이뤄지는지 그 여부를 파악할 수 있다. 이를 통해 개별 태스크의 진행에 비하여 시간의 흐름의 축에서 보았을 때 학습의 효과를 확인할 수 있다. [그림 10]은 일주일간 매일 진행된 태스크에 대해 전체 참여자의 하루 평균 소요 시간의 추이를 나타낸 그래프이다. 시도 일수의 누적에 따른 소요 시간과 표준 편차의 변화를 살펴보면, 최종적으로 1 일차와 7 일차 간 평균 소요 시간이 51 초에서 35 초로 약 31% 가량 감소했음을 확인할 수 있다 [표 4] . 두 변수 간의 영향 관계를 보기 위해 살펴본 결정 계수는 0.568 으로 유의미한 관계가 나타남을 확인할 수 있었다.

7 일동안 나타난 변화의 양상을 세부적으로 살펴보면, 먼저 첫 1 일차에서 드러난 참여자의 평균 검색 소요 시간은 51 초(SD = 63 초)로 나타났으며, 2 일차는 동일한 평균을 보였지만 그 편차에서 차이를 보였다 (SD = 56 초). 3 일차 에서는 37 초(SD = 32 초)로 시간 감축의 형태를 보였으며, 4 일차에서는 29 초(SD = 28 초)가 기록되어 1 일차에 비해 그 소요 시간이 43%가량 줄어들었음을 확인할 수 있다. 그러나 5 일차에 이르면 30 초(SD = 23 초)로 미세하게 그 평균 소요 시간이 증가하는 추세를 보이지만, 소요 시간의 편차가 줄어드는 모습이 나타난다. 이후 6 일차에서는 34 초(SD=29 초), 뒤이어 마지막 날에는 35 초(SD = 29 초)로 첫 날보다 유의미하게 검색 소요 시간이 감소했음을 확인할 수 있다.

[표 4] 태스크 수행 일자의 누적에 따른 평균 소요 시간과 표준 편차

일수	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	전체
평균	00:51	00:51	00:37	00:29	00:30	00:34	00:35	00:38
SD	01:03	00:56	00:32	00:28	00:23	00:29	00:29	00:40

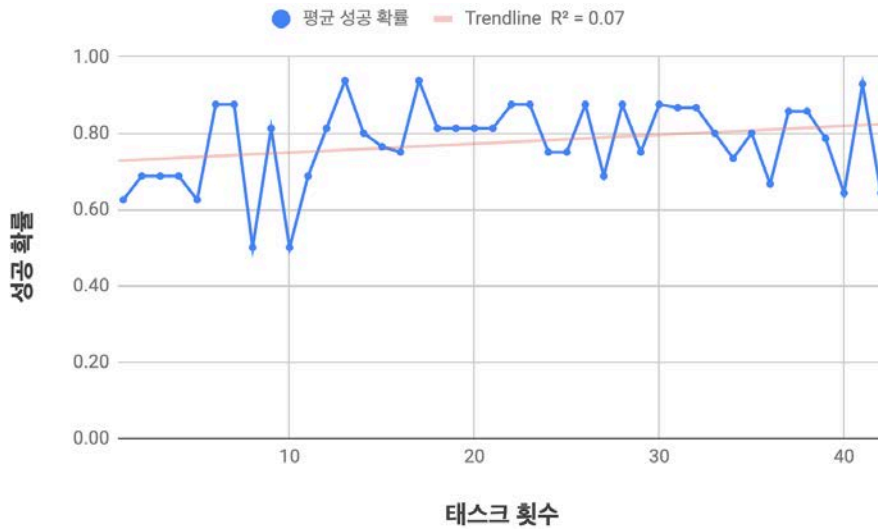


[그림 10] 수행 일자의 누적에 따른 소요 시간의 변화

본 실험에서 각 일자 별 소요 시간의 표준 편차는 사진을 검색할 때 나타나는 경험의 비일관성으로 해석할 수 있다. 효율적인 검색이 빠른 시간 내에 사용자가 원하는 대상을 찾아가는 과정을 뜻한다면 (Bergman, 2013), 검색의 소요 시간이 일정하지 않은 경우는 찾고자 하는 사진에 대해 검색에 소요되는 시간이 매번 일관되지 않게 나타났음을 유추할 수 있다. 따라서 표준 편차의 차이가 점차 줄어드는 모습은 참여자가 검색을 진행함에 따라 그 기능에 적응하고, 어떤 방식을 활용할 때 좀 더 효율적인 검색이 가능할지에 대한 학습과 적응이 이뤄지고 있음을 짐작할 수 있다.

2. 시도 횟수의 누적에 따른 실패 확률의 변화

〈연구문제 1.2〉의 검증을 위해 각 태스크에서 참여자들이 시도한 672 개의 결과에 대해 검색의 성공을 ‘1’로, 실패를 ‘0’으로 값을 변환 후, 이에 대한 평균 성공률을 계산하였다. 평균이 0 에 가까울수록 실패의 확률이 높고, 1 에 가까울수록 성공의 확률이 높은 것으로 분석되었다. 태스크의 누적 횟수에 따라 변화하는 성공 확률의 결과는 [그림 11]과 같다.

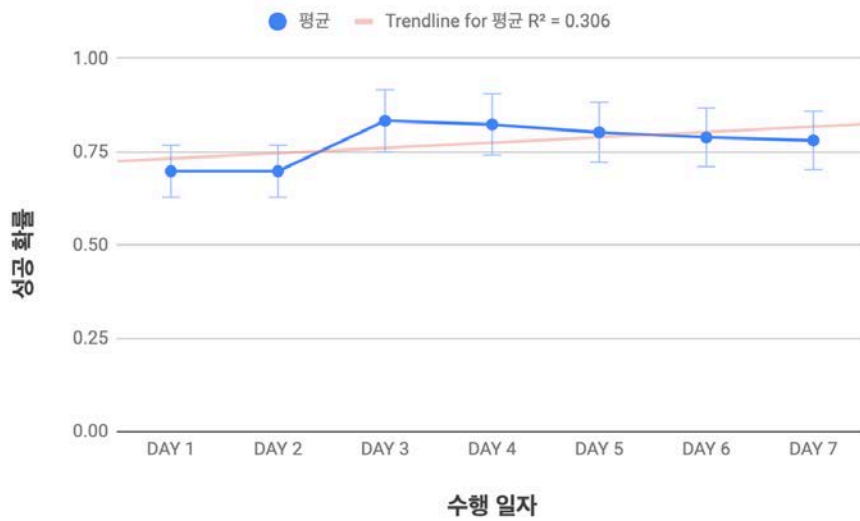


[그림 11] 태스크 누적 횟수에 따른 평균 성공 확률의 변화

〈연구문제 1.1〉의 결과와 유사하게 성공 확률이 높고 낮음의 폭이 큰 불안정한 형태를 보이고 있지만, 그 추이는 미세하게 점차 증가하고 있는 것으로 나타나고 있다. 그러나, 개별 태스크의 변화를 통해 실패 확률의 변화의 추이를 확인하기에는 그 차이를 판단하기 어렵다 ($R^2=0.07$, $p>0.05$). 따라서 〈연구 문제 1.1〉과 같이 수행 일자의 축에서 성공 확률의 변화 추이를 살펴볼 필요가 있다 [그림 12]. 〈연구문제 1.1〉에서 확인한 바와 같이 3 일자의 평균 성공 확률은 0.83($SD = 0.37$)로 가장 높게 나온 후, 마지막 일자에서는 줄어든 모습을 보인다. 그러나, 태스크 수행의 첫날($M = 0.70$, $SD = 0.46$)에 비해 마지막 날($M = 0.78$, $SD = 0.42$)의 평균 성공 확률이 약 11% 상승했음을 확인할 수 있다. 이에 따라 일주일간의 태스크 수행 기간 동안 검색의 실패 확률은 점차 줄어들었음을 유추해 볼 수 있다 [표 5].

[표 6] 태스크 수행 일자의 누적에 따른 평균 성공 확률과 표준 편차

일수	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	전체
평균	0.70	0.70	0.83	0.82	0.80	0.79	0.78	0.78
SD	0.46	0.46	0.37	0.38	0.40	0.41	0.42	0.42

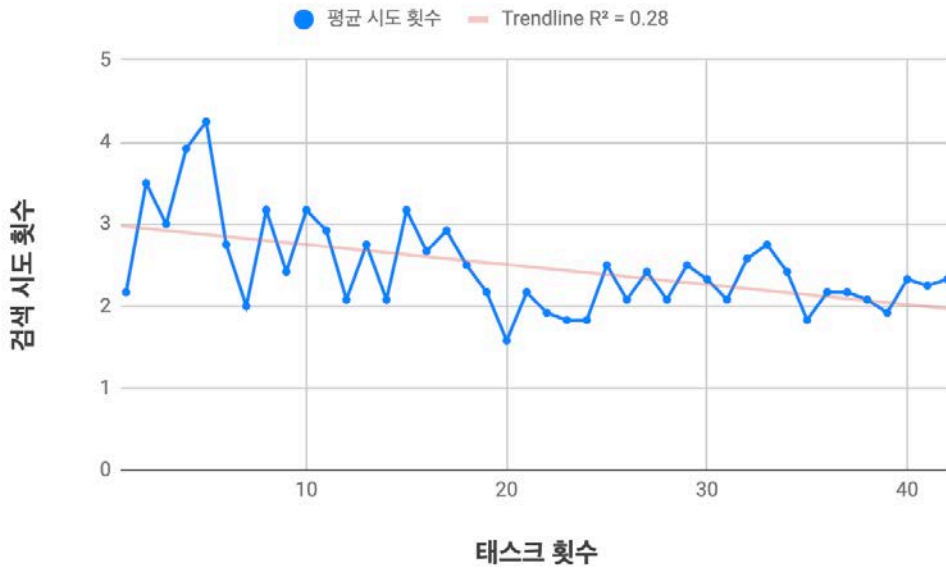


[그림 12] 수행 일자의 누적에 따른 평균 성공 확률의 변화

그러나, <연구문제 1.1>에서 볼 수 있었듯 개인 사진첩 내부에 있는 사진의 유형과 검색 대상이 되는 사진으로 인해 성공적인 검색이 일정하게 나타나지 않았고, 이로 인해 수행 일자의 누적에 따라 성공 확률이 크게 증가하는데 부정적인 영향을 미친 것으로 유추해볼 수 있다.

3. 시도 횟수의 누적에 따른 검색 시도 횟수의 변화

<연구문제 1.3>은 각 태스크에서 참여자들이 챗봇을 통해 기재한 각 태스크에서의 검색 시도 횟수 값의 변화 추이를 통해 그 결과를 살펴본다. [그림 13]에서는 태스크의 누적 횟수에 따라 변화하는 평균 검색 시도 횟수를 나열한 결과를 확인할 수 있다. 태스크의 누적 횟수에 따른 평균 검색 시도 횟수 또한 앞서 살펴본 검색 소요 시간과 평균 성공 확률의 변화 추이와 비슷한 형태를 보였다. 큰 차이로 감소하는 모습은 아니지만, 미세하게 줄어드는 모습을 확인할 수 있었다 ($R^2=0.28$, $p=0.000$).

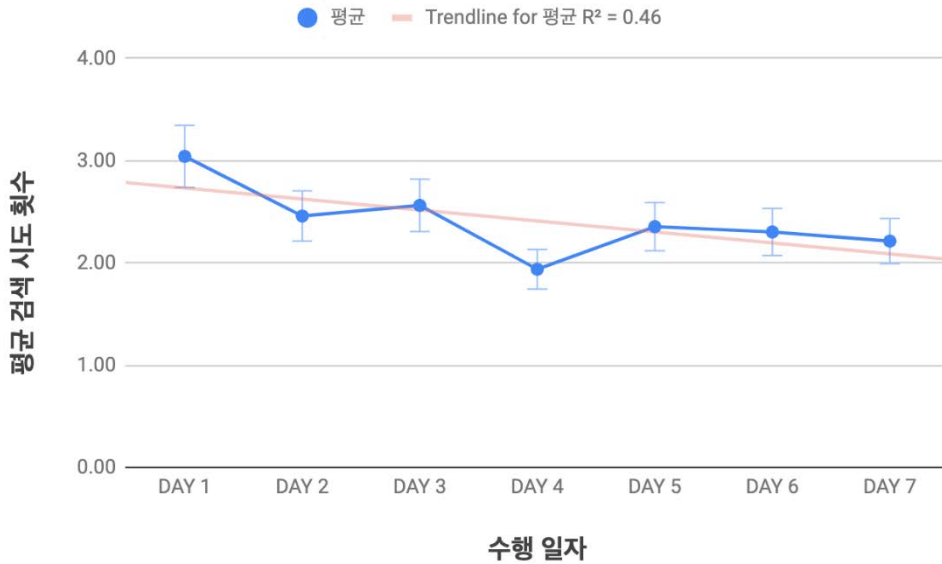


[그림 13] 태스크의 누적 횟수에 따른 평균 검색 시도 횟수의 변화

시간의 누적에 따른 검색 시도 횟수의 차이를 확인하기 위해 앞서 살펴본 바와 같이 수행 일에 따라 평균 검색 시도 횟수를 도출하였다 [표 7]. 태스크 수행의 첫날은 평균적으로 3.04 회($SD = 2.41$)의 검색 시도 횟수를 보였지만, 마지막 날에 가서는 평균 2.20 회($SD = 1.46$)으로 약 28%의 평균 횟수의 감소를 확인할 수 있다. 나아가 [그림 14]에서도 수행 일자의 누적에 따른 평균 검색 시도 횟수의 변화의 추세를 확인할 수 있다. 이를 통해 개별 태스크 횟수의 누적에 비해 보다 유효한 감소의 추이가 나타나는 모습을 확인할 수 있었다 ($R^2=0.46$).

[표 7] 태스크 수행 일자의 누적에 따른 평균 검색 시도 횟수와 표준 편차

일수	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	전체
평균	3.04	2.46	2.65	1.94	2.35	2.30	2.20	2.41
SD	2.41	1.47	1.67	1.17	1.47	1.54	1.46	1.70



[그림 14] 수행 일자의 누적에 따른 평균 검색 시도 횟수의 변화

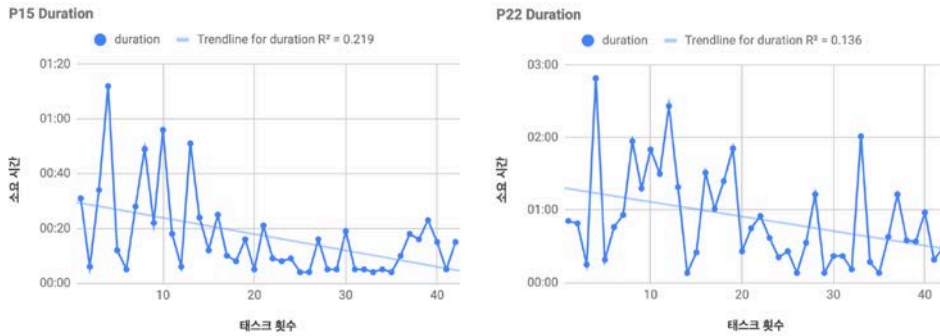
제 2 절 사용자 별 학습의 형태와 그 특성

본 절에서는 〈연구문제 2〉의 결과를 알아보기 위해 개별 사용자의 시도 횟수의 누적에 따른 소요 시간의 변화 양상을 살펴본 후, 이를 크게 학습 효과가 증대되는 경우, 반대로 학습의 효과가 떨어지는 경우, 그리고 학습의 효과가 나타나지 않는 경우로 나누어 본다.

1. 시간이 지남에 따라 학습 효과가 증가하는 경우

머신 러닝을 활용한 사진 찾기의 행동에서 학습의 모습이 드러나는지 검증한 〈연구문제 1〉의 결과에 따라, 대부분의 참여자는 검색 기능에 대해 학습과 적응이 일어나는 모습을 보였다. 실험의 참여자 총 16 명 중 12 명 (75%)의 참여자가 태스크의 누적에 따라 소요 시간이 감소하는 모습을 보였다 (P01, P02, P05, P10, P13, P14, P15, P17, P18, P20, P21, P23) [그림 15]. 학습 효과가 증가하는 경우의 평균 성공 확률은 0.81(SD = 0.48)로, 전체 참여자의 평균 성공 확률에 비해 보다 높은 확률을 보이고 있으며,

평균 검색 시도 횟수는 2.37(SD = 0.48)로 전체 참여자의 평균 검색 시도 횟수에 비해서도 더 많은 시도를 했음을 확인할 수 있다.



[그림 15] 검색 소요 시간이 감소하는 경우의 예 (좌: P15, 우: P17)

이러한 결과를 통해 여러 번의 사진 검색을 시도함으로써, 검색이 잘 되는 경험을 쌓고 구글 포토가 잘 인식하는 검색어에 대한 자신만의 전략을 쌓아나간 후 이를 이후의 검색에서도 활용하는 것으로 유추해볼 수 있다. 실제로 사후 설문에서 기재된 답변에서도 제공된 전략을 사용하여 긍정적인 검색 경험을 쌓고, 이를 통해 점차 검색이 쉬워졌다는 의견을 들을 수 있었다.

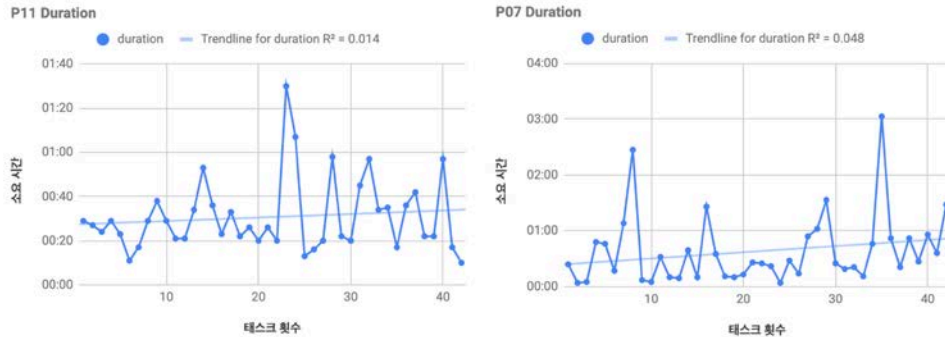
좀 더 쉽게 느껴진 것 같긴 하네요. 정확한 명칭 사용하고 사진 속 인물의 성별이나 나라 혹은 도시 이름인 음식 이름을 넣으면 잘 검색되더라고요. (P10, 사후 설문)

쉽게 느껴지고, 전략이 더 발전한 것 같다. 특히 콤마 기능을 잘 활용하면 구체적인 결과를 이끌어낼 수 있다. (P21, 사후 설문)

2. 시간이 지남에 따라 학습 효과가 감소하는 경우

총 16 명의 참여자 중, 3 명의 참가자(18.75%)는 태스크의 수행 횟수가 누적되어도 검색에 걸리는 시간이 점차 증가하는 추세를 보였다 (P06, P07,

P11) [그림 16]. <연구문제 1>에서 검증 했듯, 전반적인 사용의 행태에서는 검색에 대한 학습과 적응이 일어나는 모습이 보인 반면, 그와는 반대로 학습 효과가 나타나지 않는 참여자에 대해서는 그 행태를 주목해 봄 직하다.



[그림 17] 검색 소요 시간이 증가하는 경우의 예 (좌: P07, 우: P11)

학습 효과가 감소하는 경우, 평균 성공 확률은 0.86($SD = 0.48$)로 전체 참여자보다 더 높은 성공률을 보였다. 반면, 평균 시도 횟수는 2.06 ($SD = 0.48$)로 전체의 평균보다 낮은 모습을 나타내었다. 즉, 적은 검색어로 더 높은 검색 성공률을 보였지만, 시간이 지날수록 이러한 형태가 학습에는 도움이 되지 않았음을 알 수 있다. 그 이유를 유추하기 위해 매일 태스크 완료 후 작성한 사후 설문지와 전체 태스크의 완료 후 작성한 설문의 답변을 참고하였다.

짧에 있는 텍스트도 인식할 수 있는 줄 알았는데 텍스트라고 검색 후 수많은 사진 중에 직접 찾아야 했다. 그냥 찾는 거랑 별다른 차이가 없었다. (P06, 첫 번째 날 후기)

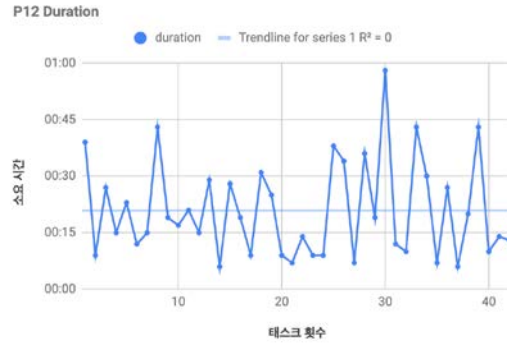
내가 원하는 검색어가 바로 나오지 않아 여러 번 검색해야해서 귀찮았음. 그래서 그냥 사진이 나올만한 검색어를 사용해서 결과가 나오면 그 중에서 찾음, (P11, 사후 설문)

이를 통해 학습의 효과가 감소하는 경우는 검색어를 여러 번 사용하여 정확한 사진을 찾을 수 있도록 시도하기보다, 처음에 검색의 전략으로 자신이 찾는 사진과 비슷한 사진이 나올법한 검색어를 설정하고, 이후 스크롤을 통해 둘러보기를 활용하는 모습을 유추할 수 있었다. 해당 방식을 사용한 참여자는 평균 소요 시간이 점차 늘어나는 결과를 보이는데, 이는 사용자의 사진첩 내부에 검색 대상이 되는 사진과 유사한 사진이 얼마나 많은지에 따라 검색에 소요되는 시간이 상이하게 나타나기도 한다. 즉, 검색을 효율적으로 할 수 있는 전략 형성에는 다다르지 못하고 단순히 수많은 사진 중에서 필터링하여 사진을 다시 찾게 되는 상황이 발생하는 것이다.

3. 시간이 지남에 따라 학습 효과가 나타나지 않는 경우

비록 16 명 중 한 명의 적은 비율(6.25%)이지만, 학습의 효과를 파악하기 어려운 참가자(P12)의 형태도 확인할 수 있었다 [그림 18]. 이 경우 평균 성공 확률은 0.12 (SD= 0.33)으로 매우 낮은 모습을 보였으며, 평균 시도 횟수는 2.71 (SD = 1.51)로 전체 참여자의 평균 시도 횟수보다 더 많은 검색을 시도하는 모습을 보였다. 이는 수많은 시도에도 불구하고 검색의 성공 경험이 매우 적었음을 뜻한다. 그에 대한 이유를 사후 설문을 통해 확인할 수 있었는데, 과거의 검색 경험이 이후의 검색어를 형성하고 학습하는데 영향을 미치는 모습을 확인할 수 있었다. 결국 반복되는 실패에 어떤 검색어가 효율적인 검색어로 사용될 수 있는지 전략의 형성이 불가능했고, 결과적으로는 매번 실패를 반복하다 검색 결과에 대한 신뢰도 떨어지게 되는 것이다.

*과거 사용한 키워드나 비슷한 종류의 키워드는 쳐도 없을 걸 알아서
더 쉽게 찾는 사진이 없다고 답하게 되어요
(P12, 다섯 번째 날 후기)*



[그림 18] 학습 추이를 확인하기 어려운 경우(P12)의 예

제 3 절 제공된 전략의 활용과 변형

본 절에서는 <연구문제 3>에 대해 답하기 위해 참여자가 실험 기간 동안 어떠한 전략을 활용했는지, 학습이 이뤄짐에 따라 팀의 활용에는 변화가 일어나는지, 그리고 검색의 경험이 누적되며 자신만의 전략을 형성하는지, 혹은 전략을 어떻게 변형해 나가는지 살펴볼 것이다.

1. 제공되는 전략의 활용

<연구문제 3.1>의 검증을 위해 각 태스크의 수행에서 참여자가 챗봇을 통해 선택한 기존의 전략과 자신이 형성한 전략을 살펴보았다. [표 8]은 16 명의 참여자로부터 일주일 간 672 개의 태스크에서 총 사용된 전략과 그 비율에 대한 결과이다. 제공된 전략이 사용된 경우는 총 87.80%으로, 대부분의 경우 참여자는 실험에서 제공된 전략을 사용하는 모습을 보였다. 그 중에서도 특히 가장 많이 활용된 전략은 ‘정확한 명칭 사용하기’ (44.35%)로 전략의 사용 중 절반에 가까운 모습을 보였다. 다음으로 많이 활용된 전략은 ‘,(콤마) 사용하기’(16.37%)였으며, 화면의 색상과 같이 ‘화면에서 드러나는 검색어 활용하기’(13.99%), ‘인물(여자, 남자) 검색어 활용하기’(7.89%) 전략이 뒤를 이었다. 제공된 전략 중, 가장 적게 사용된 전략은 ‘분류에서 검색어 참고 하기’(2.38%) 였으며, 그 다음으로 적게 사용된 전략은 ‘이벤트로 검색하기’(2.83%)로 나타났다.

[표 8] 전체 태스크에서 사용된 전략과 그 비율

제공된 전략	사용된 개수	비율
, (콤마) 사용하기	110	16.37%
인물(여자, 남자) 사용하기	53	7.89%
정확한 명칭 사용하기 (구체적인 단어 / 상위 개념)	298	44.35%
화면에 드러나는 검색어 활용하기	94	13.99%
이벤트로 검색하기	19	2.83%
분류에서 검색어 참고하기	16	2.38%
나만의 전략 사용	82	12.20%
제공된 전략 사용	593	87.80%
전체	672	100%

가장 많이 활용된 전략인 ‘정확한 명칭 사용하기’는 구글 포토가 가지는 분류 체계에 대한 인식을 형성하기 위해 작성되었다. 따라서 크게 구체적인 단어를 활용하기와 상위 개념의 단어를 사용하기로 구분할 수 있다. 그러나 실제 실험에서는 두가지 경우를 분리하지 않고 검색어를 변형할 때 사용된 전략으로 선택하도록 하였다. 이후 분석에서도 두가지 경우를 분리하지 않고 전체 전략의 활용만을 살펴보았는데, 이는 수집된 검색어의 데이터가 검색 로그를 통해 얻어진 것이 아닌 사용자의 자기 보고에 의해 기록된 데이터인 이유로 그 답변의 정확도에 대한 근거가 부족했기 때문이다.

‘정확한 명칭 사용하기’의 전략이 가장 많이 사용된 이유에 대해서는 검색어의 재조정에서 일어나는 변형 전략으로 활용한 경우 선택했음을 유추해볼 수 있다. 구체적인 하위어나 상위 개념의 단어를 질의어로 변형해나가며 시도하는 모습은 현재 구글 포토 내에서 가지는 라벨의 분류 체계에 대해 파악하기 위한 시도로 해석할 수 있다. 전체 데이터 중 검색 실패의 경우 147 개 중 61 개(전체 실패의 경우 중에서 41.22%)에 해당하는

전략도 ‘정확한 명칭 사용하기’에서 나타났다는 점에 주목해볼 직 하다. 이는 검색어 변형의 시행 착오를 겪는 와중에 사용자가 구글 포토의 분류와 일치하는 검색어를 찾으려는 시도에서 나타난 실행착오의 횟수가 많았음을 나타낸다.

그 외에도 ‘나만의 전략’을 제외한 나머지 전략에서 사용된 개수 대비 실패의 개수의 비율이 20%를 웃도는 모습을 확인할 수 있었다 [표 9]. 여기에서 검색 경험을 통해 형성된 사용자의 전략이 효율적인 검색을 위해 형성되었고, 실제로 그 실패 비율 또한 적게 나타났다. 즉, 형성된 전략은 객체 인식을 활용한 사진의 검색에서 어려움을 극복하고 효율적인 검색에 다다르는데 유효하게 사용될 수 있다.

[표 9] 사용된 전략 중 검색의 실패 비율

제공된 전략	실패 개수	사용된 전략 대비 실패 비율
, (coma) 사용하기	31	28.18%
인물(여자, 남자) 사용하기	16	30.19%
정확한 명칭 사용하기 (구체적인 단어 / 상위 개념)	61	20.47%
화면에 드러나는 검색어 활용하기	27	28.72%
이벤트로 검색하기	4	21.05%
분류에서 검색어 참고하기	3	18.75%
나만의 전략 사용	6	7.32%
전체	672	

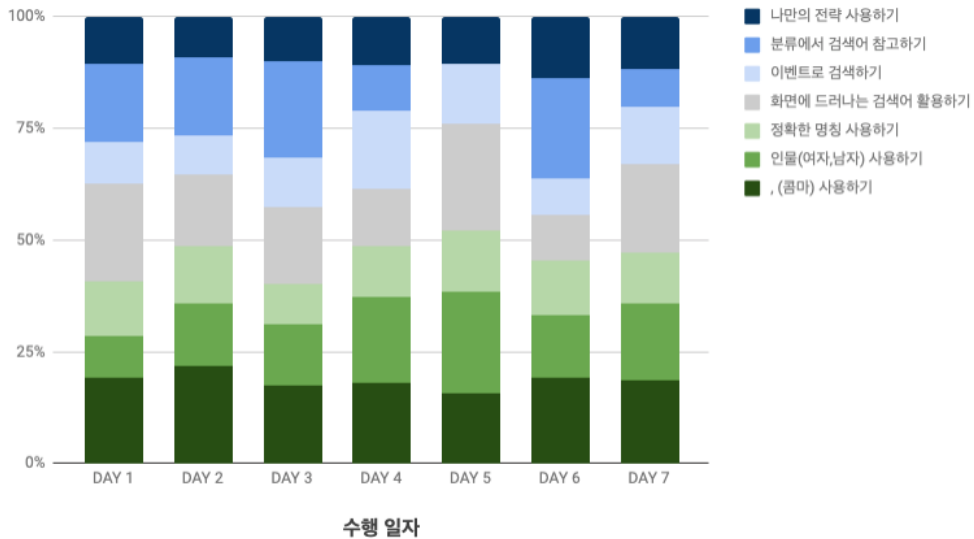
제공된 전략 중 잘 사용되지 않은 ‘분류에서 검색어 참고하기’는 구글 포토가 자체적으로 갖고 있는 분류의 단위를 사용자로 하여금 익숙해질 수 있도록 노출하기 위해 제공되었다. 그러나 참여자는 실제 검색의 경험을 통해 구글 포토의 검색 가능 범위를 탐색하는 모습을 보였다.

2. 시도 횟수의 누적에 따른 전략의 사용

〈연구문제 3.2〉의 검증을 위해 먼저 태스크 시도 횟수와 수행 일자에 따른 전략의 사용 비율 변화 양상을 살펴보았다. 그러나, 시도 횟수가 누적됨에 따라 사용 전략이 변화하는 모습은 데이터를 통해 나타나지 않았다. 수행 일자의 축에서 그 변화를 살펴보았을 때 [그림 18] 제공된 전략 중 가장 유의미한 증가를 보인 전략은 ‘인물(여자, 남자) 사용하기’로 나타났는데, 첫 번째 날 ($M = 2.00$)에 비해 마지막 날($M = 2.69$)의 전략 사용 비율이 34.5% 증가했다. 그 외 사용된 전략에는 시도 횟수가 누적됨에 따라 유의미한 증가의 모습이 나타나지 않았다. 그 이유로는 검색 대상이 된 사진에 따라 활용 가능한 검색어와 전략이 다르기 때문에 그 추이에 큰 변화가 나타나지 않은 것으로 해석해볼 수 있다. 세번째로 많이 활용되었던 ‘(코마) 활용하기’의 경우 첫번째 날($M = 4.13$)에 비해 마지막 날($M = 2.93$)의 활용이 30% 감소한 것으로 나타났다. 해당 전략의 경우, 사전 조사를 통해 다른 사용자가 유용하게 생각한 전략을 일반화하여 제공한 것인데, 기본적인 검색 기능에 대한 인지를 위해 제공하였다. 따라서 초반에 사진 검색 시 가능한 기능을 익힌 후에는 이를 적극적으로 활용하지 않고, 다른 전략을 사용하려는 모습이 보이는 것으로 해석할 수 있다.

나만의 전략의 경우, 수행 일자에 따른 평균 사용 횟수의 변화 추이는 아래 [표 10]과 같다. 표에서 확인할 수 있듯, 시간이 지남에 따라 사용자가 형성하는 전략이 점차 형성되었다고 말하기는 어렵다. ‘나만의 전략’의 경우 매일 비슷한 수준으로 나타났는데, 되려 첫번째 날에서 마지막 날로 갈수록 그 사용 빈도가 미세하게 떨어지는 것으로 나타났다. 이는 초반에 시도한 전략에 대해 그 가능성을 탐구하고, 검색이 잘되는 전략을 점차 확립해 나간 것으로 해석해볼 수 있다. 즉, 초반에 나만의 전략을 여럿 시도해본 후 그 중에서 자신의 사진첩의 특성과 기존의 검색 경험에 따라 가장 유효한 전략을 수립해 나가는 것이다. 이에 따라 점차 전략이 증가하는 모습보다는 수립해 나가는 모습이 나오는 것으로 유추해볼 수 있다. 이는 〈연구문제

3.1>에서 확인했듯, 나만의 전략을 활용했을 때 그 검색의 실패 비중이 줄어들음을 통해 유추해볼 수 있다.



[그림 18] 수행 일자에 따른 전략 사용의 변화

[표 10] 시도 횟수의 누적에 따른 ‘나만의 전략’ 활용

일수	DAY1	DAY2	DAY3	DAY4	DAY5	DAY6	DAY7	전체
평균	2.67	0.84	1.17	2.83	1.50	2.00	1.50	1.79
SD	1.03	1.17	0.75	0.41	0.55	1.26	0.84	1.09

3. 나만의 전략 형성

일주일 간 실험을 진행하며 전체 참여자가 직접 형성하여 사용한 전략은 총 82 개 (12.20%)로 나타났다. <연구문제 3.3>의 검증을 위해 새로 등장한 전략을 수집하고, 근거 이론을 바탕으로 한 개방 코딩을 진행하였다. 결과로 수집된 ‘나만의 전략’ 중 검색에 실패한 경우 6 번을 제외한 76 회에 대해 두 명의 연구자가 새로 생겨난 전략 중 검색에 성공한 경우의 전략을 분류한 후 개별 전략에서 나타나는 특성 및 개념을 군집화 하였다 [표 11].

[표 11] 참여자가 형성한 전략과 그 비율

제작된 전략	사용된 개수	비율
행정구역명의 활용	30	39.47%
자동 분류된 인물 활용	18	23.68%
건축물의 활용	10	13.61%
구글 포토에 맞춘 단어 활용	7	9.21%
특정 날짜의 활용	3	3.95%
저장 형태의 활용	3	3.95%
계절의 활용	2	2.63%
표정의 활용	1	1.32%
소품의 활용	1	1.32%
연락처의 활용	1	1.32%
전체	76	100%

참여자가 형성한 전략 중 가장 많이 활용된 개인 전략은 ‘행정 구역 명의 활용’ (39.47%)로 나타났다. 그 다음으로는 구글 포토가 자동으로 분류한 인물에 이름을 붙여 검색어로 활용하는 ‘자동 분류된 인물 활용’(23.68%)이 뒤를 이었으며, 세 번째로 많이 활용된 전략은 ‘건축물의 활용’(13.61%)였다. 그 중 ‘행정 구역 명의 활용’과 ‘건축물의 활용’의 경우, 사람이 사진을 찾을 때 기본적으로 사용하게 되는 정보가 검색어로 드러났음을 보여준다. 사진을 되찾을 때 가장 많이 활용하는 정보는 어디서 찍었는지, 언제 찍었는지와 같은 사진 촬영 당시의 맥락 정보이다 (Whittaker, 2010; Niederee et al., 2015). 더불어 스마트폰에서 촬영된 사진이 GPS 정보를 같이 저장하여 가장 정확하게 그 결과를 제공할 수 있는 방법이기 때문에 지리 정보가 많이 활용된 것으로 보인다. 다만 저장된 위치 기록과 동일한 텍스트를 활용할 때에만 원하는 결과를 얻을 수 있으므로, 장소를 나타내는 여러 표현 중 가장 정확도가 높은 행정 구역명이 사용자가 찾아낸 전략으로

활용되는 것이다. 정확한 지역명을 사용하지 않는 경우에는 그 결과가 만족스럽지 않은 경우가 존재 했지만, 이런 경우에는 행정 구역명으로 검색어를 조정하는 모습을 보였다.

홍대라고 하면 안나오는데 마포구라고 하면 나와요 (P01, 다섯 번째 날 후기)

더 좋은 팁은 고등학교 때 친구들이랑 교실에서 찍은 사진을 찾으려고 할 때 '신명여고'로 검색하면 안 나오니까, 신명여고가 위치해 있는 '인천광역시 남동구'로 검색하면 잘 나왔습니다 (P07, 사후 설문)

반면, 정확한 장소명을 사용하지 않을 경우에는 결과가 나오지 않고, 적합한 대체 단어가 떠오르지 않아 아예 다른 검색어로 활용하는 경우도 있었다. 이를 통해 장소를 통해 사진을 검색할 경우에는 위에 기재된 바와 같이 '행정 구역명'을 활용하는 것이 중요하게 작용한 것으로 해석될 수 있다. '건축물의 활용'은 주로 '덕수궁, 하남 스타필드, 코엑스'와 같이 장소를 건물의 이름이 대표할 때 사용되었다.

호수 공원 사진을 못 찾아서 결국 친구 얼굴로 검색했음 (P06, 여섯 번째 날 후기)

다음으로 많이 활용된 개인의 전략은 '자동 분류된 인물의 활용'이었다. 이는 구글 포토의 사용 횟수가 늘어나며 기능을 탐색하고, 그 와중에 시스템이 자동으로 분류해놓은 인물에 이름을 지정한 후 이를 검색어로 활용하는 모습이었다. 객체 인식이 이름 부르는 방식을 깨닫게 되는 경우는 학습된 분류에 라벨이 지정되었을 때이다. 그러나 라벨이 지정되기 전까지는 단순히 비슷한 특성을 가진 이미지를 한데 묶은데 불과하다. 따라서 구글 포토를 이용하던 도중, 사진첩 내 비슷한 얼굴로 묶어주는 자동 인물 분류를 활용하거나 가까운 사람 혹은 자주 찾는 대상의 경우에는

이름까지 붙이는 모습을 확인할 수 있었다. 이에 더해 연예인이나 유명인사의 경우에는 이름을 사용자가 지정하지 않아도 미리 학습되어 있으면 좋겠다는 목소리도 나타났다.

동일 인물끼리 사진 그룹핑 되어있어 몹시 편리하네요. 나중에 남자친구나 가족은 이름 붙여놔더니 훨씬 쉬워졌어요.
(P23, 일곱 번째 날 후기)

같이 찍은 사람을 구글 포토 얼굴 인식으로 저장해뒀을 때가 제일 잘 찾아진다. (P01, 사후 설문)

유명인사는 인식해서 검색하면 나오게 해주세요
(P16, 일곱 번째 날 후기)

사용자가 기재한 데이터 상에서 많이 나타나지는 않았지만, ‘구글 포토에 맞춘 단어의 활용’(9.21%)은 앞선 검색의 경험을 통해 구글 포토의 객체 인식 기능이 화면내에서 잘 잡아낼 수 있는 단어를 유추한 후, 그에 맞춰 작성하는 전략이었다. 이러한 모습은 초기에 검색이 가능한 단어와 가능하지 않은 단어를 탐색 이후 검색 경험이 쌓였을 때 나타나는 것으로 보였다.

검색하는데 엄청 포괄적인 단어를 먼저 쓰게되는 것 같습니다. 괜히 구체적인 단어 썼다가 안나오면 더 귀찮아질까봐서, 실제 사용할 때는 포괄적인 단어 쓰고 스크롤바로 찾아쓰는 방식으로 사용하게 될거같네요. (P17, 세 번째 날 후기)

단체로 검색하니깐 어떤 단체사진은 나오는데 가족사진은 안나와요. 그룹이나 여러명으로 검색하면 아무것도 안나와요. 팬케이크로 검색하면 팬케이크 사진이 안나와요. 케이크 하면 케이크 사진은 나오는데... (P01, 두 번째 날 후기)

다음으로는 데이터 상에서는 드러나지 않았지만, 매일 참여자가 남긴 검색 후기와 사후 설문을 통해 밝힌 사용 전략의 예시이다. 먼저 아직 현지화가 되지 않아 나타나는 검색의 어려움에 대처하기 위한 전략을 확인할 수 있다. 구글 포토의 분류 기능은 아직까지 영문을 기준으로 한다. 이는 서비스의 개발이 해외에서 진행 될 뿐 더러, 기본적인 모델의 학습은 영어로 진행되고 있기 때문이다. 이에 따라 ‘간장게장’과 같은 우리나라에만 있는 음식이나, 특정한 지명의 경우에는 검색어를 입력하더라도 적합한 결과를 얻지 못하게 된다. 전자의 경우는 구글 포토가 학습한 데이터 세트에 없는 사진이거나, 화면에서의 객체는 학습했지만 적합한 개념의 라벨을 학습하지 못한 경우로 유추해볼 수 있다. 따라서 사용자들은 일주일간의 사용 기간 중 한글을 사용하는 경우보다 영어를 직접 활용하여 검색했을 때 더 정확한 결과가 나오는 경험했을 때, 이후 태스크부터 영어를 적극적으로 활용하는 모습을 보였다.

Salmon 은 되는데 연어는 왜 안되는지 모르겠음.

(P06, 두 번째 날 후기)

오션월드 사진을 검색할 때 지역명 홍천으로 검색하거나 ocean world 로 검색하면 뜨는데, 한글로 오션월드 라고 하면 안떠요. 한글 인식 기능이 조금 더 부족한 것 같습니다. (P07, 네번째 날 후기)

특별한 음식들은 음식명으로 검색할 때 안 나오는 것 같아요. ex. 간장게장 (P07, 여섯 번째 날 후기)

영어가 짱입니다. 사실 구글포토는 백업만 해놓고 안썼는데 영어로 쓰다보니 조금씩 편해지네요 굿굿 (P17, 여섯 번째 날 후기)

장소와 연관지어 그 안에 있을 법한 사물을 활용하는 경우도 나타났다. 이 경우에는 먼저 키워드를 장소로 활용한 후, 그 결과가 너무 많아 범위를

좁히고자 시도할 때 활용되었다. 이처럼, 장소와 사물, 혹은 장소와 사람과 같이 각기 다른 종류의 키워드를 혼합하여 사용할 때 가장 효율적인 검색이 일어나는 모습을 확인할 수 있었다. 이를 직접적으로 사후 설문에서 요구하는 목소리도 있었다.

*장소에 있던 세밀한 키워드를 찾기! 공원일 때는 자전거나 벤치!
해변은 모래! (P20, 사후 설문)*

사진이 장소를 구체적으로 기억하면 좋겠어요. 예를 들면 상수동에서 찍은 사진, 혹은 OOO 가게에서 찍은 사진 이런 식으로요. ‘어디에서 누구와 찍은 어떤 사진’ 여기에 답을 다는 식으로 검색할 수 있는 도구가 있으면 좋겠다는 생각이 드네요 (P23, 사후 설문)

사후 설문을 통해 팁을 활용하기에 어려운 경우에 대한 예시도 수집할 수 있었다. 여기에는 객체 인식 기술 자체의 미흡함으로 인해 생기는 어려움을 확인할 수 있었다. 팁을 활용하기 어려운 경우는 대표적으로 객체 인식의 기능에서 인식이 정확지 않을 때 그 불편함을 가장 많이 토로했다. 인식의 우선순위가 없이 배경에 있는 사물을 인식한다던지, 특징적인 객체가 없는 일상적인 사진은 검색어 형성이 어려운 점 등에서 객체 인식의 특성을 확인할 수 있었다.

*음식이나 건물, 자연물처럼 특정 대상을 찍은 사진은 찾기 쉬운데
일상적인 셀카나 사진을 찾기는 어려운 것 같아요. 인식이 백퍼센트
되지 않기도 하고 검색어를 잡고 검색하기도 애매한 것 같습니다.
(P06, 사후 설문)*

더불어 스마트폰 사진첩에 있는 사진이기 때문에 나타나는 어려운 점도 확인할 수 있었다. 자신의 사진첩에 있지만, 직접 촬영하여 저장된 사진인지 혹은 다운 받은 사진인지에 대해 확실지 않다는 점이였다. 각각의 경우에는 활용 가능한 검색어가 달라지는데, 머리 속에 떠오른 사진이 직접 촬영인지

혹은 공유 받은 사진인지 헷갈릴 경우에는 어떤 검색어가 가능한지 알 수 없게 되는 것이다.

카톡으로 사진을 공유를 많이 받다보니, 이게 내가 찍은 사진인지 받은 사진인지, 받아서 내가 저장해서 나한테 있는 사진인지 아닌지가 매우 헷갈림. (P21, 여섯 번째 날 후기)

내 사진인지 남이 찍어줘서 내가 저장한 적이 있는건지 헷갈린다. 사진이 여기저기 혼재 되니까 머릿속에 있는데 그게 진짜 구글 포토나 클라우드에 있는지 모를 때가 있다. (P21, 일곱 번째 날 후기)

제 6 장 연구 논의

제 1 절 객체 인식을 활용한 사진 검색에서 나타나는 기기와 인간의 차이

본 연구에서는 <연구문제 1>을 통해 객체 인식 기술이 활용된 사진 검색에서 일주일 간의 실험 기간에 걸쳐 학습의 양상이 나타남을 밝혔다. 이어서 <연구문제 2>를 통해 참여자를 학습이 일어나는 경우와 되려 학습이 어려운 경우, 효과가 없는 경우로 분류하고, 학습이 다르게 나타나는 이유에 대한 분석을 진행하였다. <연구문제 3>에서는 기기의 특성을 사용자에게 전략으로 제공했을 때 수용되는 전략과 변형되거나 새로 생겨나는 전략에 대해 살펴보았다. 세 가지 연구문제를 통해 살펴봐야 할 점은, 사진을 되찾을 때 활용되는 객체 인식에 대해 어떤 특성이 유용하게 활용되며, 어려움은 어떤 지점에서 나타나는지에 대한 것이다. 이는 앞서 살펴본 선행 연구에 따라 크게 세 가지 지점에서 논의해볼 수 있다. 먼저 객체 인식이란, 컴퓨터가 화면 상에 나타난 표면을 읽어내고 그 안에 포함되어있는 객체를 감지하여 적합한 이름을 부여하는 것이다. 이 과정에서 차이가 나는 지점을 크게 세 단계로 나눌 수 있다. 첫 번째로 인식의 대상, 즉 스마트폰의 사진의 특성은 컴퓨터 비전이 표면을 읽어내는데 어려움을 야기한다. 두번째로 대상을 인지하는 방식에서 인간과 기기는 차이를 가지게 된다. 마지막으로 인식한 객체에 대해 이름 지어 부르는 단계에서 차이가 발생할 수 있다.

먼저 인식하는 대상의 특성은 구글 포토 내에서 컴퓨터 비전이 인지하는 대상이 스마트폰 내에서 촬영된 ‘개인 사진’의 모음이라는 점에서 발생한다. 관련 연구에서 살펴보았듯, 사진기로 촬영한 사진은 촬영자의 관점을 내포하고 있다는 특성을 지닌다. 이는, 객체 인식이 머신러닝을 통해 모델을 학습하는 사진과는 달리 실제 사용자의 사진첩에서 그 기술이 적용될 때 읽어내야 하는 사진과 큰 차이를 가진다는 데에서 문제가 발생한다. 사용자

각각의 독특한 관점이 적용된 사용자의 스마트폰 사진첩은 모델이 학습한 사진과는 달리 해석하기 어려운 사진들의 누적이기 때문에 분류의 정확도에 차이가 나타나는 것이다. 나아가 촬영된 사진은 객체인식이 읽어내기 쉬운 특정한 대상으로 한 사진이 될 수도 있지만, 분위기 혹은 풍경, 접사와 같이 다양한 형태의 화면으로 나타날 수도 있다. 특히 사진의 경우, 대상이 명확하지 않으면 촬영한 사람도 이름짓기 어려운 경우가 있었다. 일상적인 사진이거나, 추상적인 사진, 혹은 어딘가에 비친 사진과 같이 그 대상이 한번에 드러나지 않는 사진이 주로 그 대상이 되었다.

*음식이나 건물, 자연물처럼 특정 대상을 찍은 사진은 찾기 쉬운데
일상적인 셀카나 사진을 찾기는 어려운 것 같아요. 인식이 백퍼센트
되지 않기도 하고 검색어를 잡고 검색하기도 애매한 것 같습니다.
(P06, 사후 설문)*

*일상 사진처럼 딱히 특징이 없는 경우엔 검색어로 찾기가 어려운 것
같네요 (P07, 일곱 번째 날 후기)*

위와 같은 이유는 모델이 학습된 바와 달리 실제 적용될 때 낮은 인식률을 보이는 이유가 될 수 있다. 이러한 지점에서 확인할 수 있듯, 현재의 객체 인식은 대상을 명확하게 알아볼 수 있거나, 비교적 뚜렷한 개념을 가지는 사진에 대해 학습이 진행되고 있다. 이를 미뤄볼 때, 개인 사진과 같이 실 생활에 적용되어 복잡한 형태를 가진 대상에 대한 학습도 이뤄져야 할 필요가 있다.

두 번째로 인식의 차이에서는 인간과 기계가 화면 내에서 주시하는 대상의 우선순위가 다르기 때문에 나타나는 것으로 해석해볼 수 있다. 주로 나타나는 어려움은 종종 객체 인식 기술이 과도하게 세밀한 부분까지 잡아내어 너무 많은 사진이 결과로 나타난다는 점이었다. 앞서 살펴본 관련 연구에 따르면, 사람은 화면을 인식할 때 가장 먼저 두드러지는 대상을 인식하고, 공간 내에서 대상과의 거리 및 관계를 인지한 후 그 외의 것들에

대해 인식하는 것과 같은 순서를 따른다. 그러나, 현재의 객체 인식은 인식의 순서 없이 화면 내에서 객관적으로 읽히는 객체를 모두 결과로 제안하고 있다. 이에 따라 사람이 화면에서 보고싶은 것과 기기가 화면에서 읽어내는 것 사이에 차이가 발생하는 것이다. 나아가 아직 기기는 사물 혹은 화면에 찍힌 사람의 미묘한 표정이나 관계를 읽어내기 어렵다. 실제로 이러한 지점은 참여자가 사후 설문을 통해 지적한 어려움의 지점과 맞닿아 있었다.

사진에 너무 작은 부분도 인식되서 사진이 너무 많이 나와요 (ex. 한강에서 찍은 사진 찾을 때 한강, 잔디, 독섬 키워드로는 안나오더니 자전거 찾을 때 저 멀리 작게 나온 자전거 인식해서 나오네요ㅠ) (P20, 다섯 번째 날 후기)

같은 키워드의 사진에서 미묘하게 포인트(표정 등)가 다른 사진을 찾기 어려움. (P13, 네 번째 날 후기)

마지막으로 인식한 객체를 호명하는 단위의 차이가 있을 수 있다. 사회문화적으로 분류되어 개인이 인식하는 범위 내에서 이름지어 부르는 사람과 달리, 화면 내에서 나타나는 특징을 통해 객체를 분류하고, 해당 특징을 가진 라벨을 찾아 짝짓는 기기 사이에는 분명 어떠한 간극이 존재한다. 이 간극이란 사람이 모르는 개념에 대해 기기가 알고 있거나, 기기의 세세한 분류 단위를 한데 묶어 사람은 하나로 이름지어 부르는 경우가 있을 수 있다. 일례로 ‘미국 지빠귀’를 그 분류 자체로 이름 부르는 것이 아닌, ‘새’로 지칭하는 것과 같은 것이다. 인간은 분류하지만 기기가 학습이 되지 않아 분류하지 못하는 경우도 있다. ‘물회,’ ‘간장게장’과 같이 인간에게는 한눈에 보고 당연하게 인식하는 대상이 아직 기기에게는 기록되지 않아 라벨을 붙이기 어려운 것이다. 마지막으로 기기가 가지는 분류보다 사람이 인식하고 이름지어부르는 분류가 더 촘촘한 경우도 존재한다. 같은 케이크라도 ‘팬케이크’, ‘크레이프 케이크’, ‘조각 케이크’ 등 그 분류가 다양하지만, 기기는 아직 세부적인 단위를 인지하지 못한다.

그러나 현재는 사람이 모든 경우의 수를 직접 경험하며 현재 구글 포토가 인식 가능한 범위에 대해 탐색해나가야 한다. 이러한 지점에서 사람이 사용하는 분류와 기기의 분류의 차이점을 좀 더 심도있게 다룰 필요가 있다.

제 7 장 결론 및 연구의 의의

제 1 절 연구의 요약

객체 인식 기술이 인간의 능력을 뛰어넘을 만큼 발전하고난 후, 스마트폰 사진첩에 적용되어 개인 사진의 검색이 가능해졌다. 그러나 이를 적극적으로 활용하기에는 그 분류의 결과가 아직 사람이 보기에 어색하고, 부정확한 점이 있다. 더불어 사람이 사용하는 검색어에 적합한 결과가 나오기 어렵다. 그러나, 혼잡한 사진첩에서 새로운 기술을 활용하여 관리 및 되찾기(Retrieval)을 도우려는 시도는 필요하다. 기술이 접목되어 이제 막 사용되기 시작한 기능에 대해 사람들이 부정적인 인식을 갖게 될 경우, 이후의 사용에도 좋지 않은 영향을 미칠 가능성이 있다. 이러한 맥락에서 본 연구는 스마트폰 사진첩에서의 사진 검색의 문제 지점을 찾고, 개선 가능성에 대한 탐색을 위해 시작되었다. 특히 처음으로 사진 검색 기능을 활용하는 사람들에게서 학습의 모습이 보이는지, 기기의 특성을 이해할 수 있도록 전략을 제공할 경우 이를 수용하는지, 이 두가지 관점에서 연구를 진행하였다.

이를 위해 선행 연구에서는 먼저 사진 검색을 가능하게 만든 배경인 컴퓨터 비전의 발전과 현재 사용되고 있는 이미지 검색에서의 객체 인식의 활용, 그리고 객체 인식을 통한 이미지 검색이 가지는 철학적 함의에 대해 살펴보았다. 두 번째로는 현재 객체 인식에서 생기는 근본적인 문제인 기기와 사람 사이의 ‘인식 방법의 차이’에 대해 알아보았다. 이를 기반으로 추후 연구에서 활용되는 사진 검색의 전략을 작성할 수 있었다. 마지막으로 검색의 대상이 되는 사진에 대한 다양한 논의와 그 의미를 읽고, 그것이 현재 스마트폰에서 활용될 때 나타난 변화와 활용 가능성에 대해 살펴보았다.

뒤이어 진행된 실험은 사용자들이 객체 인식이 접목된 사진 검색 기능을 활용하며 학습의 모습을 보이는지, 제공된 전략을 얼마나 수용하는지 알아보기 위해 진행되었다. 이를 위해 구글 포토의 검색 기능을 처음 사용해보는 16명의 사람들에게 일주일간 매일 여섯번의 사진 검색 태스크를 제공하였다. 매일 여섯개의 사진을 찾으며 참여자들은 태스크에 대해 떠오른 사진, 검색 시도 횟수, 사용한 검색어, 사용한 전략을 기재하였다. 하루의 태스크가 마무리되면 그 날의 검색 태스크에 대해 어려웠던 부분을 짧게 기술하였다. 이렇게 일주일간 모인 태스크 별 데이터는 총 672 개였다. 매일 검색이 끝난 후 어려웠던 지점 혹은 불편했던 부분에 대한 간단한 설문을 작성하고, 일주일 간의 실험이 끝난 후에는 사후 설문을 통해 전반적인 검색에 대한 경험을 서술하였다. 이렇게 모인 설문의 응답은 추후 데이터를 해석하고 결과의 이유를 유추하는데 활용되었다.

객체 인식을 활용한 사진 검색에서 학습의 형태가 나타나는지에 대한 질문을 제기했던 <연구문제 1>의 결과로는, 42 회의 태스크가 진행되며 전체 참여자의 평균 검색의 소요 시간이 짧아지고 실패율이 적어지는 형태가 나타났다. 즉, 객체 인식을 활용한 검색의 활용에 어느정도 적응을 한 모습을 보였다. 태스크 횟수의 누적이 아닌, 일주일의 시간 축에서 확인했을 때에도 그 소요 시간과 실패율의 감소는 동일하게 나타났다.

<연구문제 2>에서는 전반적인 학습의 모습에 반해 몇몇 사용자에게서 나타나는 개별적 행태를 확인할 수 있었다. 16 명 중 3 명의 참가자는 학습이 이뤄지지 않고 검색의 소요 시간이 증가하는 모습을 확인할 수 있었다. 그러나 초반의 성공률은 전체의 평균보다 높은 모습을 보였는데, 이는 검색어를 여러 번 사용하여 정확한 사진을 찾을 수 있도록 시도하기보다, 처음에 검색의 전략으로 자신이 찾는 사진과 비슷한 사진이 나올법한 검색어를 설정하고, 이후 스크롤을 통해 둘러보기를 활용하기 때문에 나타난 것으로 해석되었다. 또 다른 경우로는 학습이 일어나지 않은 사람이 한 명 관찰되었는데, 이 경우 반복되는 검색 실패의 경험으로 인해 이후

검색어를 활용하는 데 있어 결과에 대한 신뢰도가 매우 낮아진 것으로 해석될 수 있었다.

〈연구문제 3〉에서는 일주일동안 제공된 전략이 얼마나 활용되었는지, 참여자가 개별적으로 만든 전략은 어떻게 발전되어 갔는지를 알아보기 위해 태스크 수행에 따른 전략의 활용과 시간의 축에서 전략 활용의 변화를 살펴보았다. 이를 통해 가장 많이 활용된 전략은 ‘구체적인 단어나 상위 개념을 사용하여 정확한 명칭으로 검색어 활용하기’였다. 이는 검색에서 결과를 도출하기 위한 적합한 검색어를 찾아나가는 과정 중에 가장 많이 사용될 수 밖에 없는 전략인 것으로 유추해볼 수 있었다. 태스크 시도 횟수가 누적됨에 따라 ‘,(콤마) 활용하기’와 같은 전략은 점차 그 사용이 줄어드는 모습을 보였는데, 이는 처음에 활용 방식에 대한 학습이 진행된 후에는 다른 전략을 활용하는 것으로 옮겨간 것으로 추측해볼 수 있다. ‘나만의 전략’의 경우 일주일간의 사용의 축에서 큰 변화 없이 비슷한 비중으로 활용되는 모습을 보였다.

본 연구의 주요 발견은 객체 인식 기술이 사진 검색에 활용될 때, 그 기능에 대한 학습의 가능성을 확인했다는 점이다. 특히 그 과정에서 전략의 활용 형태와 추후 수집한 설문을 통해 객체 인식 기술과 인간의 사용 행태 사이에서 발생하는 간극의 지점을 확인한 것은 본 연구의 가장 중요한 발견이라고 할 수 있다. 더불어 사진이라는 특성을 통해 객체 인식 기술이 실생활에 활용될 때 나타나는 어색함과 부정확함에 대한 이유를 사진 찾기 태스크를 통해 확인할 수 있었다는 점에서 그 의미를 가진다.

연구의 결과와 논의에서 다룬 내용을 바탕으로 사진 검색을 위한 객체 인식의 학습 모델과 학습 데이터 세트, 그리고 검색 인터페이스 측면에까지 개선되어야 할 부분에 대한 제언이 가능하다. 먼저 학습의 측면에서는, 검색 기능을 처음 사용하는 사용자에게 현재 객체 인식의 한계와 검색을 활용하는 방법에 대한 가이드를 제시할 수 있다. 이를 활용할 경우 사진이 검색되지 않는 경우에도 직접 많은 경우의 수를 경험하는 것 보다 문제를

빠르게 해결할 수 있을 것이다. 검색 인터페이스의 측면에서는, 사용자가 검색어를 형성하기 어려울 때 어떤 검색이 가능한지 단서를 제공하는 인터페이스적 요소가 적용될 수 있을 것이다. 검색 가능한 질의어를 추천하기 위해서는 사용자가 입력한 검색어에 대해 시스템 내에서 검색 가능한 범위로 유도할 수 있도록 사용자의 질의어와 시스템의 분류간의 유의어, 상의어, 하의어를 연결짓는 시소러스(Thesaurus)가 필요할 것으로 예상된다. 이를 통해 사용자는 구글 포토의 객체 인식이 가지는 분류를 이해하게 되고, 점차 효율적인 검색어를 활용할 수 있을 것이다.

본 연구의 결과는 추후 객체 인식 기술을 스마트폰 내의 개인 사진 검색에 활용할 때, 사용자가 느끼는 어려움의 지점을 개선하는데 활용될 수 있을 것이며 추후 사람들이 사진을 자신의 방식대로 자연스럽게 효율적으로 찾는데 초석을 다지는데 기여할 수 있을 것으로 기대한다.

제 2 절 연구의 한계

본 연구는 실험 진행자와 참가자가 물리적으로 떨어진 공간에서 챗봇을 통해 원격으로 진행되었다. 실험 진행에 대한 설명을 구두로 설명하는 것이 아닌, 가이드로 제공했다는 지점에서 이유로 참가자의 실험에 대한 이해도는 차이를 가질 수 있다. 더하여 실험 진행자와 참여자가 물리적으로 같은 공간에서 조사를 진행하는 실험실 세팅과는 달리, 챗봇을 실험 조사 도구로 사용할 경우에는 참가자의 자발적인 참여와 의지가 중요해진다. 그러나 본 실험은 실험실 환경에서 실행한 조사에 비해 참가자의 성실한 대답을 유도하기 어렵다는 점에서 실험 설계에 대한 아쉬움이 남는 바이다. 그러나, 챗봇을 활용하여 알림을 제공하고, 실험에 대한 적절한 부담과 동기 부여를 통해 사용자의 참여를 유도했다는 점에서 유의미한 시도인 것으로 사료된다. 나아가 설문지 형태뿐만 아니라, 실험 조사를 챗봇을 통해 진행했다는 점에서 연구 조사 방법론으로써 챗봇의 활용 가능성을 엿볼 수 있다.

실험의 설계 측면에서 아쉬웠던 부분은 참가자로 하여금 사진을 찾는 자연스러운 상황과 목적을 따로 설정하지 않고, 특정 사진을 찾도록 유도했다는 점이다. 이는 사진을 되찾는 상황이 비교적 빈번하게 일어나지 않는다는 점에서 기인했다. 태스크 형태의 장치를 제공함으로써 실제의 검색 상황을 떠올릴 수 있도록 요청하는 것 이상의 자연스러운 환경을 형성하기에는 그 사진 되찾기의 빈도가 비교적 낮은 편이었으므로, 대안으로 인위적인 장치를 설계할 수밖에 없었다. 이에 따라 실제의 사진 검색 상황과는 차이가 있을 수 있다는 점에서 그 한계를 갖고 있다. 더불어 실험 참여자가 16 명으로, 비교적 적은 수를 모집하여 표본을 대표하기 어렵다는 점도 본 연구의 한계점으로 볼 수 있다. 이는 초기에 시작한 참가자 수에 비해 일주일간 꾸준히 태스크를 수행하고, 데이터를 입력한 참가자가 비교적 드물었기 때문에 더욱 어려운 문제로 다가왔다.

본 연구를 통해 도출된 전략은 일주일 동안 실험 진행자로부터 제공된 사진 찾기 태스크의 환경 내에서 형성된 전략이다. 이에 따라 실제 환경에서 사진을 찾을 때 자연스럽게 나타나는 전략과는 차이가 있다는 데 그 한계가 있다. 그러나, 아직 객체 인식을 활용한 사진 검색이 많은 사람들에게 사용되지 않고, 기술이 접목된 초기 시점이라는 사실을 감안할 때, 본 연구의 결과를 바탕으로 추후 자연스러운 환경에서의 사진 검색 전략을 살펴볼 필요가 있다. 덧붙여 시도 횟수뿐만이 아니라, 시간의 축에서 학습의 모습을 관찰할 때 일주일이라는 기간은 비교적 짧은 시간에 속한다. 보다 긴 기간동안 자연스러운 환경에서 사진 검색의 행태를 관찰한 후 검색 경험의 설계를 세심하게 다듬어야 할 필요가 있다. 나아가 전략 활용의 이유를 설문이 아닌 인터뷰를 통해 진행했다면, 보다 폭넓고 다양한 이유를 확인할 수 있었을 것이다.

마지막으로는 본 논문에서 객체 인식을 활용한 사진 검색 기능의 개선 지점을 파악하고, 그 개선 방안에 대한 가능성만을 제시하고 있다. 본 연구를 통해 도출된 주요 결과를 반영한 프로토타입을 구현하여 실제

검색의 효율성과 편리성을 기존의 것과 비교 분석 하였다면, 보다 더 실증적으로 강력한 연구 결과로써 제시될 수 있을 것이다.

제 3 절 연구의 시사점

본 연구의 학술적인 시사점은 다음과 같다. 먼저, 머신 러닝을 활용한 객체 인식 기술이 개인 사진의 검색에 적용되었을 때에도, 기기의 특성을 제공하여 이를 통해 학습이 일어남을 밝혀내었다. 사진 검색은 그동안 GPS 정보, 날짜 정보와 같이 사진 촬영 시 같이 기록되는 메타 데이터 혹은 사용자가 직접 붙이는 태그를 통한 검색에 집중되어왔다. 자동 주석을 통해 사진 관리와 찾기를 도우려는 시도는 있었지만, 구글 포토와 같이 다수의 사용자에게 제공되는 서비스를 통해 그 학습의 행태를 관찰하기는 어려웠다. 반면 본 연구는 기술이 실 생활에 적용되었을 때 초기의 사용자가 이를 시도하는 모습을 관찰함으로써 일주일이라는 시간의 축에서 학습의 모습을 밝혀내었다.

둘째, 개인의 사진을 객체 인식을 활용한 검색을 통해 찾는다는 데에서 검색 연구 분야의 확장에 기여하였다. 일반적인 웹에서 나타나는 이미지 검색 행동과 달리 스마트폰에서 접근할 수 있는 개인 사진을 되찾는 데에는 사용되는 검색어가 달라질 수 있으며, 특히 사용자가 직접 분류한 것이 아닌 객체 인식을 활용하는 경우에는 더 큰 차이가 드러날 수 있다. 나아가 객체 인식 연구의 분야와 이미지 검색 분야, 그리고 사진이라는 다의적인 매체를 종합하여 현재 기술이 가지는 문제점에 대한 고찰을 시도했다는 점에서 그 의의가 있다.

본 연구가 산업 기술적 측면에서 가지는 시사점은 다음과 같다. 객체 인식은 지금까지 학계에서 그 능력을 고도화하기 위한 연구의 대상이었다. 그 기술이 실제 사용자에게 제공되어 어떻게 활용되는지, 객체 인식의 특성이 사용자에게 학습이 될 수 있는지 그 여부에 대해서는 알려진 바가

부족한 상황이었다. 본 연구에서는 객체 인식 기술이 적용되었을 때 가지는 기기의 특성에 대해 사용자가 수용할 수 있는 특성과 받아들이기 어려운 특성을 구분해 내었다는데 사진 검색 시스템에서 객체 인식 기술을 어떻게 전달해야 하는지에 대한 단서가 될 수 있다.

둘째, 기기의 특성을 사용자에게 제공하여 이를 활용한 사진 검색 행동에서의 학습 가능성에 대해 탐구했다. 전략을 활용하여 학습이 진행되는 경우와 반대로 학습이 이뤄지지 않는 경우로 나눌 수 있었다. 이는 유효한 전략을 사용자에게 제공할 경우, 처음의 부정확함과 어색함을 극복할 수 있는 가능성을 보여준다. 나아가 학습이 일어나지 않는 경우에 대해서는 본 연구를 통해 유추한 이유를 통해 이를 실증적으로 검증한다면, 추후 다양한 사용자에게 대해서도 검색 경험을 증진시킬 수 있는 방안을 마련할 수 있을 것이다.

마지막으로 융합적으로는 HCI(Human Computer Interaction)의 관점에서 본 연구를 보았을 때, 기기의 특성을 전략의 형태로 사용자에게 제공하고 사용자가 검색 과정을 경험하며 전략을 수용하고 변형해나가는 과정에 집중했다는데 그 기초를 따랐다고 할 수 있다. 더하여 현재 객체 인식의 활용에 대해 인문학적 고찰을 시도하고, 사진의 문화적, 철학적 의미를 탐색함으로써 객체 인식 기술이 적용되어야 할 방향에 대한 논의를 시도했다는데 그 의의가 있다.

참고 문헌

- Barthes, R. (1981). *Camera lucida: Reflections on photography*.
조광희, 한정식(역) (1998). <카메라 루시다 : 사진에 관한 노트>. 서울: 열화당.
- Berger, J. (1980). *Understanding a photograph. Classic essays on photography*. 김현우(역) (2015). <사진의 이해>. 서울: 열화당.
- Flusser, V. (1983). *Towards a philosophy of photography*. 윤종석(역) (1999). <사진의 철학을 위하여>. 서울: 커뮤니케이션북스.
- Sontag, S. (2001). *On photography* (Vol. 48). 이재원(역) (2005). <사진에 관하여>. 서울: 도서출판 이후.
- 박병선. (2012). 스마트폰 ‘텍스트 사진’ 이용 행태와 내용 분석에 관한 연구. 석사학위논문, 서울대학교 대학원.
- 오세욱, 이소은, & 최순욱. (2017). 기계와 인간은 커뮤니케이션할 수 있는가?. *정보사회와 미디어*, 18(3), 63-96.
- 이재현. (2009). 디지털 에크프라시스. *한국언론학보*, 53(5), 244-267.
- 이재현. (2019). 객체 인식과 바르트: 폰크툼의 죽음, 미발행 간행물.
- 정광태, & 홍자인. (2008). 제품의 학습성을 평가하기 위한 학습곡선 모델의 적용. *대한인간공학회지*, 27(2), 59-65.
- 조수진. (2012). 기억처리과정의 이해.
- Bergman, O. (2013). Variables for personal information management research. In *Aslib Proceedings: New Information*

Perspectives (Vol. 65, No. 5, pp. 464–483). Emerald Group Publishing Limited.

Bhagat, P. K., & Choudhary, P. (2018). Image annotation: Then and now. *Image and Vision Computing*, 80, 1 – 23.

Broekhuijsen, M., van den Hoven, E., & Markopoulos, P. (2017). From PhotoWork to PhotoUse: exploring personal digital photo activities. *Behaviour & Information Technology*, 36(7), 754–767.

Brooks, L., Rosch, E., & Lloyd, B. B. (1978). Cognition and categorization.

Das, A., Agrawal, H., Zitnick, L., Parikh, D., & Batra, D. (2017). Human attention in visual question answering: Do humans and deep networks look at the same regions?. *Computer Vision and Image Understanding*, 163, 90–100.

Jaimes, A. (2006). Human factors in automatic image retrieval system design and evaluation. In *Internet Imaging VII* (Vol. 6061, p. 606103).

Jiang, L., Kalantidis, Y., Cao, L., Farfadi, S., Tang, J., & Hauptmann, A. G. (2017). Delving deep into personal photo and video search. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 801–810). ACM.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive psychology*, 16(2), 243–275.

Kim, S., Lee, J., & Gweon, G. (2019). Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 86). ACM.

- Kindberg, T., Spasojevic, M., Fleck, R., & Sellen, A. (2005). The ubiquitous camera: An in-depth study of camera phone use. *IEEE Pervasive Computing*, 4(2), 42 – 50.
- Kofler, C., & Lux, M. (2009). *An exploratory study on the explicitness of user intentions in digital photo retrieval*. na.
- Le, Q. V., Ranzato, M. A., Monga, R., Devin, M., Chen, K., Corrado, G. S., ... & Ng, A. Y. (2011). Building high-level features using large scale unsupervised learning. *arXiv preprint arXiv:1112.6209*.
- Mandler, J. M., & Parker, R. E. (1976). Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology*, 2, 38 – 48.
- Malik, A., Dhir, A., & Nieminen, M. (2016). Uses and gratifications of digital photo sharing on Facebook. *Telematics and Informatics*, 33(1), 129–138.
- Nielsen, J. (1994). *Usability engineering*. Elsevier.
- Niederee, C., Kanhabua, N., Gallo, F., & Logie, R. H. (2015). Forgetful digital memory: Towards brain-inspired long-term data and information management. *ACM Sigmod Record*, 44(2), 41–46.
- Nieuwenhuysen, P. (2018). Information Discovery and Images A Case Study of Google Photos. *IEEE 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services*, ETT LIS 2018, 16 – 21.
- Oliveira Serrano de Andrade, D., Fernando Maia, L., Feitosa de Figueirêdo, H., Viana, W., Trinta, F., de Souza Baptista, C., & Serrano de Andrade, O. (2018). Photo annotation: a survey. *Multimed Tools Appl*, 77, 423 – 457.

- Ordonez, V., Liu, W., Deng, J., Choi, Y., Berg, A. C., & Berg, T. L. (2016). Learning to name objects. *Communications of the ACM*, 59(3), 108–115.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211 – 252.
- Sarvas, R., & Frohlich, D. M. (2011). *From snapshots to social media—the changing picture of domestic photography*. Springer Science & Business Media.
- Van House, N. A. (2011). Personal photography, digital technologies and the uses of the visual. *Visual Studies*.
- Whittaker, S., Bergman, O., & Clough, P. (2010). Easy on that trigger dad: a study of long term family photo retrieval. *Personal and Ubiquitous Computing*, 14(1), 31–43.
- Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, 45(1), 346–362.
- Zhou, J., & Chen, F. (Eds.). (2018). *Human and Machine Learning*. Springer International Publishing.
- Zhou, W., Newsam, S., Li, C., & Shao, Z. (2017). Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval. *Remote Sensing*, 9(5), 489.

Abstract

A Study of User Experience of Object Recognition based Photo Retrieval

– On Learning of Photo Retrieval Behavior
in Google Photos –

Selki Kang

Department of Digital Contents Convergence

The Graduate School

Seoul National University

The practices of photo retrieving on personal smartphones have extended. Not only pictures are browsed by scrolling up and down, but also a picture is easily come out as a result of typing a certain keyword. The technology of object recognition has changed how people look and browse personal photos; it not only classifies similar photos, but also assigns the labels that represent the referent of the classification. For instance, Google Photos has applied the object recognition and search system to allow users to manage personal photos. Eventually, the novel use of searching photos on personal album is expected to change the aspects of how people retrieve a particular photo out of thousands of accumulated ones in their cloud system.

However, the novel technology is on its early stage and is at lack of leaving positive impression to the users. There is a gap between object recognition executed by the device and the user. When typing a query on its search bar for instance, the result is either none or countless number of results.

The purpose of this study is to identify of the points of inconvenience in smartphone photo albums using object recognition and to deliver a better photo search user experience. In this paper, previous studies and preliminary research were thoroughly reviewed and held to fully grasp understanding of the inner workings of the object recognition, and to build a general frame of how people use the personal photo search system. In the main research, six photos search tasks were designed for a week with a total of 16 participants, aged between 20s and 30s. Search strategy tips were given to the experiment participants in order to collect particular strategy when searching photos. After collecting a total of 672 search tasks and used strategies, a post-questionnaire was followed.

As a result of the analysis, the learning process of search system of the users has occurred. The study was able to identify how users learned the functions of photo search through strategy. As the number of 42 retrieval tasks, the average retrieval time of 16 users gradually decreased. As a result, the average retrieval time of the last day compared to the first day decreased by 31% from 51 seconds to 35 seconds. The average success rate of searches also rose by about 11% over 42 tasks performed each day for a week. The

average number of search attempts of participants decreased by 28%. As the experience of photo retrieval is accumulated with the strategy provided, it is confirmed that the image of the learning and the improvement of the retrieval are improved in the photo retrieval using object recognition technology.

In the learning style of individual users, 12 out of 16 participants participated in the appearance of learning, and 3 showed that learning did not take place. The other one showed no influence on learning. If learning is not or is not affected, it can be inferred that there is a difference in learning depending on which strategies are used in the initial search and how to adapt to the search function.

Finally, 44.35% of the total search strategies were applied to the most commonly used strategies: 'use the correct name by using the high-level word (abstract concept) and the low-level word (concrete concept)'. Next was followed by using '(comma)', 'using search terms that appear on the screen like color', and 'using figures (women, men)'. The strategy that users individually formed was 12.20% of the total, and there was no difference in use over time. As for the strategies that the user made, 39.47% of the strategies were 'utilization of administrative area names', followed by 'automatic classification of people' and 'utilization of buildings'. Among the strategies that have been developed, 'Word usage tailored to Google Photos' has been found to recognize the characteristics of object recognition through experiences and form a search word by predicting it. In other words, as the

experience of photo retrieval using object recognition accumulates, it shows the understanding of its characteristics.

Through the above analysis, the study has examined the point where object recognition technology is difficult for user when it is used as a search in the smartphone photo album, and added a brief suggestion on how to supplement it.

This study approached the user 's difficulty in using the object recognition technology as applied to the smartphone photo album. In addition, HCI (Human Computer Interaction) side has focused on the process of how the strategy made through the viewpoint of the device is accepted and transformed by the user. In addition, it is meaningful that the study tried to observe the interaction that occurred when the research of object recognition, which was concentrated only on improving the accuracy of recognition, was provided to actual users. Finally, it is meaningful that the study discussed the ways to utilize object recognition in order to utilize the medium of photography and sustainable use.

Keywords : Machine learning, Object Recognition, Image Retrieval, Photo retrieval

Student Number : 2017-26435